Domain-Specificity of Self-Regulated Learning Processing in Science and History

Date of submission: May 31, 2015

Abstract

As computer-based learning environments grow in prominence, so do the demands placed upon students to learn with these tools. Empirical research has shown that students who are effective at self-regulating their learning are more likely to acquire deep conceptual understanding while using these environments. However, there is a noticeable lack of research into the degree to which self-regulated learning (SRL) is domain-specific. Investigating this theoretical question about domain-specificity results in related questions about how to best capture and model SRL. To address these concerns, we randomly assigned college students to either a science or history digital library, and used think-aloud protocol (TAP) data to examine the degree to which SRL processing predicted knowledge gains, above and beyond the effects of prior knowledge. We examined multiple methods of aggregating SRL TAP data into analysis variables, to determine which would be the most predictive of learning gains, and then tested these findings using a sample from a second study. In addition, we tested whether the frequency of SRL processing differed by academic domain. We found that data-driven aggregation methods were the most effective at predicting learning gains, and that there were both intriguing similarities in SRL processing across domains (e.g., the importance of corroborating sources) as well as differences (e.g., the predictive validity of self-questioning). Our findings have implications for how to capture and model SRL processing, as well as how to foster SRL among those students who do not yet enact it effectively on their own.

Keywords: self-regulated learning; science; history; think-aloud protocols; domain-specificity; digital libraries

Domain-Specificity of Self-Regulated Learning Processing in Science and History

The idea that today's students are "digital natives" (Palfrey & Gasser, 2008; Prensky, 2001) who are naturally adept at using computers to learn is, frankly, false (Bennett, Maton, & Kervin, 2008; Kirschner & van Merriënboer, 2013; Selwyn, 2009). Unfortunately, while these students are often skilled at online browsing, gaming, and socializing, they struggle when using computer-based learning environments (CBLEs) to accomplish academic tasks (Borgman et al., 2000; Mervis, 2009; Nasah, DaCosta, Kinsell, & Seok, 2010). This disparity between society's expectations of these "digital natives" and the reality of their struggles is particularly concerning given the rapid proliferation of computers into primary, secondary, and post-secondary classrooms (Collins & Halverson, 2009). Within the scholarly literature, there is ample evidence that students' ability to enact self-regulated learning (SRL; Winne & Hadwin, 2008; Zimmerman, 2000) knowledge and skills is a key predictor of their success when learning with CBLEs (Azevedo, 2005; Greene, Bolick, & Robertson, 2010; Greene, Hutchison, Costa, & Crompton, 2012; Lajoie & Azevedo, 2006). SRL involves the active, thoughtful monitoring and control of a person's cognition, metacognition, motivation, and behavior. Numerous studies have identified SRL processes that are advantageous when learning with CBLEs, including judgments of learning and creating subgoals. However, the literature lacks empirical investigations of the degree to which SRL knowledge, skills, and behaviors are domain-specific (i.e., whether the predictive validity and utility of particular SRL processes vary across subjects, such as science, history, or literature). The existence of this gap, and the need to fill it, were stated clearly by Alexander, Dinsmore, Parkinson and Winters (2011):

Empirical research on SRL has not systematically or seriously regarded the potential effects of domains or domain-specific tasks on individuals' regulation of their cognition, motivation, or emotions...we hold that this lack of domain-specific attention within the SRL empirical research demands immediate and considerable attention if educational

researchers are going to achieve a deeper understanding of the very nature of self-regulation and its role in human learning and development. (p. 403)

Increasingly, researchers have been calling for investigations of the domain-specificity, or contextuality, of learning processes such as SRL, motivation (King & McInerney, 2014) and epistemic cognition (Greene & Yu, 2014; Sandoval, 2012). We share these scholars' sense of urgency, but caution that such investigations must be conducted in ways that accurately capture the phenomena in question.

Although there have been a few investigations of the domain-specificity of SRL (e.g., Rotgans & Schmidt, 2009), these studies have relied upon retrospective, self-report SRL instruments. The numerous concerns about the accuracy (Winne & Jamieson-Noel, 2002) and utility (Veenman, 2005; Winne & Perry, 2000) of participants' self-report data calls into question any findings based upon them. As an alternative, Greene, Dellinger and Tüysüzoğlu (2013) contended that researchers should utilize concurrent measures (Winne & Perry, 2000), such as think-aloud protocols (TAPs; Ericsson & Simon, 1993), to study SRL, and in particular to investigate the degree to which SRL depends upon the domain, task, and environment. However, concurrent measures such as TAPs pose two challenges. First, they are resource-intense in terms of data collection and preparation (Greene, Robertson, & Costa, 2011). Second, TAPs produce large amounts of data that can be challenging to analyze and understand. Therefore, the theoretical research question regarding how SRL varies, or not, by domain is matched by an equally important methodological question regarding how to conduct this work in a way that is both feasible, and accurate.

We responded to the theoretical question about the domain-specificity of SRL posed by Alexander et al. (2011) by randomly assigning college students to either a science or history digital library. Given the paucity of research on domain-specificity in SRL, multiple studies,

using both between- and within-subjects designs, are needed. In this work, we utilized a between-subjects design where each participant engaged in a single, relatively long learning task (i.e., 30 minutes), to increase the likelihood that participants would have sufficient time to go through the various phases of SRL (e.g., task definition, goal setting, studying, adaptation; Winne & Hadwin, 2008). To study the relationship between SRL and learning within each domain, and to compare participants' SRL across domains, we chose comparable learning tasks and assessments that required the acquisition of both declarative and conceptual knowledge of the content.

Our research also required the development and investigation of new methodologies for preparing and analyzing large amounts of SRL TAP data. We compared Greene and Azevedo's (2009) SRL TAP data preparation and analysis methodology to Greene and colleagues' (2013) methodology to determine which best predicted learning in each of our academic domains. Then, we conducted a partial replication of these methodologies with a different sample, to further investigate their predictive validity. Our methodological findings, as well as our findings regarding the domain-specificity of SRL that emerged from those methods, have implications for theory and future research. Our findings also have implications for the design of SRL interventions within and across academic disciplines (Bembenutty, Cleary, & Kitsantas, 2013).

2. Literature Review

2.1 Managing Learning in Computer-Based Learning Environments

CBLEs have a number of important affordances, including the ability to present vast amounts of sometimes abstract information in multiple, non-linear ways that users can control themselves (Azevedo, 2005; Jacobson & Archodidou, 2000; Mayer, 2005). For example, one of the CBLEs utilized in this study, the National Science Digital Library (NSDL), provides a

comprehensive, curated network of text, image, interactive, simulation, and video based resources related to a variety of science topics. Digital libraries share the affordances of CBLEs, while also benefitting from curators who verify the accuracy and utility of the resources. By integrating the myriad of resources typically available via CBLEs, students can develop deep understandings of content in science, history and other academic domains (Azevedo, Johnson, Chauncey & Burkett, 2010; Greene, Costa, Robertson, Pan, & Deekens, 2010; Koedinger, Aleven, Roll, & Baker, 2009).

However, researchers have also shown that students struggle to learn with CBLEs, because they have difficulty coordinating multiple representations of information into coherent understanding, as well as trouble differentiating the relevant from the salient but irrelevant information (Ainsworth, 2006; Liu & Hmelo-Silver, 2009). These tasks can overwhelm working memory capacity and lead to cognitive overload, specifically high mental load (Baddeley, 2001; Gerjets, Scheiter, & Schuh, 2008; Moos, 2013; van Merriënboer, & Sweller, 2005).

2.2 Self-Regulated Learning

Researchers have shown that students who enact effective SRL processes are far more likely to overcome mental load and acquire deep conceptual understanding from CBLEs than those who do not (Azevedo, 2005; Azevedo, Guthrie, & Seibert, 2004; Azevedo, Moos, Johnson, & Chauncey, 2010; Azevedo, Johnson, Chauncey & Burkett, 2010; Greene, Costa, et al., 2010; Greene et al., 2012; Greene, Moos, & Azevedo, 2011; Jacobson & Archodidou, 2000; Koedinger et al., 2009; Lajoie & Azevedo, 2006). Although multiple frameworks exist, there is general consensus that SRL is a process where learners actively set goals for learning, and then monitor and control their behavior towards achieving those identified learning goals (Pintrich, 2000; Winne & Hadwin, 2008; Zimmerman, 2000). Winne and Hadwin's (2008) model of SRL has

four phases: task definition, planning and goal setting, studying tactics, and adaptation. During task definition, learners develop an understanding of what the learning task will require of them. Learners then create benchmarks or goals to guide and monitor their learning progress during planning and goal setting. Learners engage the use of various tactics (e.g., taking notes, summarizing, or corroborating sources) in order to achieve their desired learning goals during the third phase. Finally, in the adaptation phase learners reflect on their learning decisions, progress, and success to make critical decisions about what actions to take next. Additionally, self-regulating learners will assess the effectiveness of the various strategies and behaviors that they enacted, to inform their own knowledge about how to be a more adaptive learner.

Winne and Hadwin (2008) asserted that as learners move through these phases, they utilize metacognitive monitoring and control to make learning decisions. The metacognitive monitoring process occurs when learners compare learning products or outcomes to their goals. Depending on the progress indicated by this comparison, learners may make adaptations, such as creating new goals, changing strategies, or even adjusting their knowledge about the effectiveness of strategies they have used.

2.3 Domain Specificity in Self-Regulated Learning

Despite the large body of SRL scholarship, researchers have conducted few empirical investigations into how SRL processing might vary across learning domains, contexts or environments (e.g., how planning, strategy use and monitoring may, or may not, differ across science and history CBLEs; Alexander et al., 2011). This leaves a particularly interesting gap in the literature, as a number of SRL researchers (Alexander, 1995; Boekaerts, 1999; Greene et al., 2013; Poitras & Lajoie, 2013) have asserted that SRL processing likely has significant domain-specific aspects. For example, Poitras and Lajoie (2013) have presented a conceptual argument

that incorporates findings from empirical SRL studies into a theoretical model of how students develop a deep conceptual understanding of historical events and their causes, as opposed to solely the declarative knowledge on which such understanding is built. They argued that, over the course of learning in history, successful students use increasingly complex and domain-specific strategies (e.g., corroboration and self-questioning), and adopt more nuanced learning goals (e.g., identifying system-based causes versus singular or person-based causes). These complex and domain-specific strategies are needed to achieve deep conceptual understanding of historical phenomena, the multiple interactive causes of these phenomena, and how historical phenomena most often arise from systems of humans in interaction, rather than from solitary actors (Carretero & Lee, 2014; Wineburg, 2001).

Poitras and Lajoie (2013) have made a compelling conceptual argument regarding domain-specific aspects of SRL in history; however, their claims require empirical investigation. In one of the few examples of empirical research on the domain-specificity of SRL, Rotgans and Schmidt (2009) found no notable differences in motivation or use of SRL strategies across math, history, and English domains. However, Rotgans and Schmidt relied on the *Motivated Strategies for Learning Questionnaire* (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1991) self-report instrument of SRL for their analysis, and this type of instrument has been subject to criticisms. These criticisms have led to research into the optimal ways to capture and model SRL processing.

2.4 Capturing and Measuring Self-Regulated Learning

2.4.1 Self-report instruments. Self-report instruments (e.g., surveys) are the predominant means of capturing SRL (Duncan & McKeachie, 2005), but they remain a controversial one (Winne & Perry, 2000). Many researchers have asserted that learners are typically poor at accurately reporting the amount, type, and nature of the learning activities in

which they engage, due to the idiosyncratic and even non-conscious (i.e. implicit) nature of SRL processing (Veenman, 2007; Winne & Jamieson-Noel, 2002). Such instruments tend to characterize SRL as an attribute of the person, rather than a series of events enacted in context (Winne & Perry, 2000). Further, some researchers have argued that the validity of students' responses to the MSLQ, and similar self-report instruments, is threatened by social desirability (Bartels & Magnum-Jackson, 2009). Finally, in most cases these surveys are administered after learning has ended (i.e., they are "offline" measures, Winne & Perry, 2000). It is unclear how well students recall their SRL behavior after the fact, calling into question whether retrospective judgments are accurate representations of processing during specific learning tasks.

2.4.2 Think-aloud protocol measures. Validity concerns have led some researchers to argue for the use of concurrent online measures that model SRL as a series of events (Winne & Perry, 2000), such as think aloud protocols (TAP; Ericsson & Simon, 1993; Greene, Robertson, et al., 2011), to capture the SRL processes that learners utilize as they progress through a learning task. TAPs involve individuals verbalizing all the thoughts they have while working on a learning activity. Such data (e.g., "I didn't understand that") can be used as indicators of SRL processing (i.e., judgment of learning). Research has shown that TAPs do not interfere with cognitive processes, including SRL, when individuals are asked only to report and not explain what they are thinking (Bannert & Mengelkamp, 2008; Ericsson, 2006; Ericsson & Simon, 1993; Veenman, Elshout, & Groen, 1993). On the other hand, numerous published studies using TAPs have revealed relations between SRL processing and learning outcomes (e.g., Azevedo, Moos, Greene, Winters, & Cromley, 2008; Greene & Azevedo, 2007; Greene et al., 2012). Thus, TAPs have notable advantages over survey instruments, including being concurrent rather than retrospective measures. In addition, they are open-ended measures, allowing for participants to

verbalize whatever they wish, rather than having to respond to researcher-generated lists of SRL processes or beliefs, such as can be found in Likert-type items in survey instruments (Greene, Robertson, et al., 2011).

2.4.3 Challenges of capturing and modeling self-regulated learning with think-aloud protocols. Despite their potential, using TAPs to collect data on SRL poses serious challenges. TAP data collection and preparation (i.e., transcribing and coding) are resource-intensive tasks that average four to six person-hours of work per 30-minute learning activity (Greene, Robertson, et al., 2011). 30min of TAP data can include 30 to over 100 codeable verbalizations, each of which must be accurately coded as an SRL process from schemes that can include 40 to even 70 different SRL processes (e.g., Azevedo et al., 2008; Greene, Yu, & Copeland, 2014). Aside from the resource challenges related to the time it takes to transform SRL TAPs into usable data, this process also results in a large number of variables. Each one captures how many times participants enacted a particular SRL process. With 40 to 70 such variables, it can be difficult to collect a sufficiently large sample to have the power to conduct quantitative analyses of how different SRL variables predict learning. Some kind of data aggregation is needed (Greene et al., 2013).

Greene and Azevedo (2009) developed a methodology to address these power concerns. The codes applied to SRL TAPs are indicators of what they called micro-level SRL processing. Micro-level SRL processes are specific activities or cognitions that learners verbalize, such as taking notes, experiencing a feeling of knowing, or self-questioning. Greene and Azevedo conceptualized a two-tiered system where micro-level SRL codes are nested within broader macro-level categories (e.g., planning, monitoring, and strategy use) that align with the key aspects of Winne and Hadwin's (2008) model of SRL. As an example, there are a number of

ways that learners can monitor during SRL, including judgments of learning, feelings of knowing, and monitoring progress toward goals. Each of these specific, micro-level SRL processes, is a kind of monitoring, and aggregating the frequency of each micro-level SRL process into a single macro-level variable results in significant data reduction (see Appendix A).

Aggregating the micro-level SRL data into macro-level SRL variables (e.g., summing the frequencies of use of all strategy micro-level codes into a macro-level strategy use variable), results in a relatively small set (i.e., three to five) of variables that are theoretically important yet also reasonable to study in terms of power concerns in quantitative analysis. Aside from increasing statistical power by reducing many micro-level codes into a small set of macro-level SRL variables, such aggregation also may control for interindividual, non-essential differences in the specific micro-level SRL processes that learners enact (i.e., it may not matter whether a participant summarizes or uses a mnemonic, what matters is that a retention strategy is used; Greene & Azevedo, 2009).

Data reduction in SRL research is becoming more common. For example, Greene and Azevedo (2009) found that aggregating micro-level SRL processes to the macro-level resulted in the finding that the frequency of monitoring, in general, predicted learning gains above and beyond prior knowledge. Such a finding might have otherwise been missed if analyses were conducted on the individual micro-level SRL processes. However, as discussed by Greene and colleagues (2013), SRL processing, and consequently what micro-level SRL behaviors should be captured and analyzed, may necessarily differ by academic domain and context. Therefore, in any particular domain, context, or task, it may be difficult, a priori, to determine which micro-level SRL processes to aggregate, and which may be irrelevant in that particular context.

2.5 Capturing and Modeling Domain-Specific Self-Regulated Learning Processing

Our two research foci result from the need to investigate the domain-specificity of SRL, and the affordances and challenges of doing so using TAPs. To address our interest in domain-specificity of SRL, we first identified the specific processes and strategies that learners use in particular academic domains. There is literature suggesting that certain SRL processes can be classified as more or less adaptive regardless of academic domain. For example, high-level strategies like knowledge elaboration are more consistently associated with comprehension and learning than low-level strategies like rereading (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Schraw, 2006). On the other hand, there are almost certainly strategies that are specific to particular academic domains (Greene & Yu, 2014).

2.5.1 Science learning. Deep conceptual understanding of science topics is difficult to convey (Chi, 2005; Chinn & Brewer, 1993; Hmelo-Silver & Azevedo, 2006). Many science topics, such as the circulatory system or the phase changes of matter, are comprised of multiple levels of information that are related in intricate, non-obvious, or even counterintuitive ways (Azevedo, 2005; Jacobson & Wilensky, 2006). Multimedia representations of science content, such as CBLEs, afford opportunities for educators to represent phenomena that are otherwise invisible or inaccessible (Jacobson & Wilensky, 2006). Unfortunately, many curricula and assessments continue to focus upon memorization of the parts of systems at the expense of a deep conceptual understanding of the system as a whole, or why normative models of those systems are better than alternative models (Duschl, Schweingruber, & Shouse, 2007; National Research Council, 2005). Effective science learners must use sophisticated strategies, such as the control of variables, to not only understand the declarative aspects of systems, but also coordinate among those declarative aspects, to create dynamic conceptual understanding of scientific phenomena (Hmelo-Silver & Azevedo, 2006; Osborne, Ratcliffe, Collins, Millar, &

Duschl, 2003).

2.5.2 History learning. Learning with primary sources engages students in the active process of historical inquiry. Historical inquiry is a dynamic process that requires students to identify, evaluate, contextualize, and interpret multiple sources to construct a historical narrative (Hicks & Doolittle, 2009; Lee, 2002). According to Burton (2005), historical CBLEs enable teachers and students “to use computers to do history in ways impossible without the computer” (p. 207). The advent of CBLEs has increased teachers’ ability to use historical inquiry in the classroom because of the ease of access to digitized primary sources (Bolick, 2006; Lee, 2002; Mason et al., 2000). However, the introduction of historical inquiry into the classroom requires a variety of learning processes (e.g., sourcing, incorporation of multiple, sometimes conflicting perspectives into a multifaceted schemata) that are often foreign to history students. Given research showing that learning in science and history can involve different processes or strategies (Greene & Yu, 2014), when investigating SRL across these academic domains, researchers must allow for the possibility that what is predictive in one domain (e.g., control of variables strategy in science) may not be predictive, or even relevant, in another (i.e., history).

2.5.3 SRL data aggregation within domains. Our second research focus, on methodologies for aggregating SRL TAP data, derives from our interest in domain-specificity. The potential differences in SRL processing across domains opens up a possibility first acknowledged by Greene and colleagues (2013): that aggregating all micro-level SRL process data into macro-level SRL variables may not be sufficiently nuanced. It may be the case that only particular micro-level SRL processes are relevant or necessary for a particular domain, task, or DL. If that is the case, then summing those relevant micro-level SRL processes with other,

non-relevant, or even unhelpful, SRL processes may decrease the predictive validity of the subsequent macro-level SRL variable.

Greene and colleagues (2013) suggested that instead of aggregating all micro-level SRL data into macro-level SRL variables (i.e., full aggregation, Greene & Azevedo, 2009), a data-driven aggregation methodology might be necessary. This more iterative process involves first identifying those micro-level SRL processes that are found to be predictive of learning in a particular domain, and aggregating only these processes into macro-level SRL variables. For example, the frequency with which learners use each micro-level SRL process can be correlated with learning outcomes to identify only those SRL processes most predictive of learning. Then, only those most predictive micro-level SRL processes could be aggregated into macro-level SRL variables, rather than all micro-level SRL processes within a category (i.e., full aggregation). Greene and colleagues' suggestion to use data-driven aggregation methods is supported by previous study findings (Azevedo, Greene, & Moos, 2007; Greene & Azevedo, 2007) where particular micro-level codes (e.g., knowledge elaboration) were found to be more predictive of positive learning outcomes than other micro-level codes (e.g., controlling context) from the same macro-level group (e.g., strategy use). An initial study (Greene, Yu, et al., 2014) using this data-driven aggregation method revealed that it predicted learning gains more effectively than full aggregation (i.e., Greene & Azevedo, 2009) methods.

This more adaptive and iterative “data-driven” analysis method suggested by Greene and colleagues (2013) affords both benefits and challenges that must be addressed. Using data to determine which micro-level SRL processes to aggregate allows researchers to avoid including irrelevant or unhelpful processes in the macro-level SRL variables. On the other hand, the post-hoc nature of the data-driven aggregation method raises concerns regarding validity and the

dangers of data mining. Such concerns can be addressed by cross-replicating findings across multiple studies conducted with different participants in similar conditions and learning environments (Makel & Plucker, 2014). If concurrent measures of SRL, such as TAPs, are to continue to demonstrate predictive validity and utility, they must effectively scale to both larger samples and coding schemes. Therefore, it is critical that researchers investigate new ways of aggregating micro-level SRL data into macro-level SRL variables that have strong predictive validity. Further, investigations of the domain-specificity of SRL require coding schemes that can capture both the general SRL processes common across domains, as well as those processes that are unique. Our study is the first to address both of these current needs in SRL research.

2.6 Learning Tasks and Measures Across Domains

Finally, one last challenge when studying domain-specificity of SRL is identifying learning tasks and outcome measures that are relatively commensurate across academic domains. In our case, it is difficult to imagine a science learning task that could also be used as an authentic learning task in history, outside of investigations of socio-scientific issues in the public understanding of science (Bromme & Goldman, 2014). Instead, such research requires tasks that are commensurate in terms of the kinds of knowledge (e.g., declarative, conceptual) required to complete them successfully. Our tasks, outlined in section 3.1, required learners to explore authentic questions in each domain and use high-level learning strategies to achieve the necessary declarative and conceptual knowledge. We also endeavored to make the assessments commensurate, by including both multiple choice items that tapped declarative knowledge as well as open-ended questions that required conceptual understanding.

2.7 Overview of the current study

Clearly, there is a need for research into how SRL processing does and does not differ across academic domains. However, that research must be done in ways that allow for defensible inferences regarding SRL processing (i.e., use of online SRL data collection techniques like TAPs), while also capturing domain-based differences in the utility of strategies and other SRL processes that learners enact. In essence, research is needed that allows for an understanding of how learners' SRL processing differs within and across domains, and how those differences, or lack thereof, predict learning outcomes.

One major objective of this study was to examine how SRL processing predicted the acquisition of knowledge in two CBLEs from different academic domains (i.e., science and history), and then examine how that processing differed across those CBLEs. To best meet this objective, we had to also investigate the predictive validity of full and data-driven aggregation techniques for combining micro-level SRL TAP data into macro-level SRL variables. In Study 1 we compared the full and data-driven aggregation methods to create macro-level SRL variables that had maximum predictive validity in the science and history CBLEs, respectively. In Study 2, we conducted a partial replication to evaluate whether the macro-level SRL variables created in Study 1 would have similar predictive validity with a different sample.

Study 1 Research Questions:

- 1.1) Does participants' knowledge of the science or history topic increase from pretest to posttest, after using the CBLE to learn?
- 1.2) To what degree does participants' SRL processing, using the full aggregation method, predict posttest knowledge, above and beyond the effect of pretest knowledge?
- 1.3) To what degree does participants' SRL processing, using the data-driven aggregation method, predict posttest knowledge, above and beyond the effect of pretest knowledge?

1.4) Does the frequency of use of micro-level SRL processing differ, on average, across the two digital libraries?

Study 2 Research Questions:

2.1) Does participants' knowledge of the science or history topic improve from pretest to posttest, after using the CBLE to learn?

2.2) To what degree does participants' macro-level SRL processing, modeled after the findings for Study 1 Research Question 1.2, predict posttest knowledge, above and beyond the effect of pretest knowledge?

3. General Method and Design Overview

Two studies were conducted to address our research questions. In this section, we describe the methodology that is common across both studies. In Sections 4 and 5, we present the methods unique to each study, as well as the results and discussion appropriate to each study. All differences in participants, procedures, and methods are reported in the description of Study 2.

3.1 Procedures Common to Both Studies

We used a procedure similar to that implemented by Azevedo, Greene, and colleagues (e.g., Azevedo & Cromley, 2004; Greene, Bolick, et al., 2010; Greene, Costa, et al., 2010). Each session was held in a research laboratory, with one participant and one researcher present. First, the researcher randomly assigned participants to either a science or history CBLE. Next, participants were given 20 minutes to complete a pretest. The participants were then given a brief overview of the CBLE. Subsequently, the researcher read the learning task aloud. The learning task for the history CBLE was to understand the origin and controversies regarding the construction of the Blue Ridge Parkway. The learning task for the science CBLE was to

understand the phase change process as substances move from solid to liquid and then to a gaseous state (see Appendix B).

Next, the researcher explained the TAP process in detail, which involved instructing the participants to verbalize everything they were thinking or reading during the learning task as they navigated through the CBLE. Fox, Ericsson, and Best (2011) found that asking participants to verbalize, but not explain, their thinking did not lead to reactivity (i.e., changes to their cognition). Therefore, we stressed to participants that they simply had to say what they were thinking, without comment or reflection upon that thinking. Participants were fitted with an audio microphone to facilitate recording, and the researcher sat well behind the participant, to decrease the likelihood that the participant would feel compelled to either talk to, or monitor the reaction of, the researcher. The participants practiced the TAP process while exploring a website unrelated to the learning task. The researcher provided any necessary feedback to help the participant properly enact the TAP, and once it was clear the participant was comfortable with the procedure, and able to execute it successfully, the practice was ended.

The participants were then given 30 minutes to complete the learning task. The participants could take notes, though notes could not be used on the posttest. Participants were both audio and video recorded. If the participant was silent for more than two seconds, the researcher prompted the participant with the statement: "Please say what you are thinking." This was the only prompt used, and it was designed to be as minimally intrusive as possible. Ericsson (2006; Fox et al., 2011) has studied TAP protocols, and found evidence that asking participants to say what they are thinking does not have appreciable effects upon cognition, whereas asking participants to explain their thinking does (e.g., "Explain what you are doing"). We were careful to ensure that researchers only prompted with the "say what you are thinking" language, to

minimize the chance of TAP reactivity (Fox et al., 2011). The majority of participants required no more than one or two instances of prompting over the course of the 30min learning task.

Verbal time prompts were given to each participant when there were 20, 10, 5, and 2 minutes remaining. Upon completion of the 30 minutes of learning time, the participants had 20 minutes to complete a posttest for which no notes or instructional materials could be used. Before exiting the research laboratory, participants were asked to complete a demographic questionnaire and then engaged in a debriefing regarding the study. Participants were also asked to refrain from sharing the details of the procedure with other potential participants.

3.2 CBLEs Used in Both Studies

The participants were asked to explore one of two CBLEs, which in this study were DLs. DLs are curated collections of multimedia representations (e.g., pictures, videos, text, maps, etc.) that allow users to explore topics of interest. Both DLs that we chose were developed and managed by public institutions, and were freely available to the public. The history DL consisted of a new digital collection housed within the award-winning *Documenting the American South* (*DocSouth*) archive titled “Driving Through Time: The Digital Blue Ridge Parkway in North Carolina” (<http://docsouth.unc.edu/blueridgeparkway>). This digital collection included a rich collection of resources (e.g., text, photographs, newspaper articles) focused on detailing the history of the Blue Ridgeway Parkway. The science DL was comprised of resources (e.g., text, pictures, simulations) related to helping individuals understand the mechanics of phase changes in states of matter. This digital collection was housed within the National Science Foundation’s *Computational Science Education Reference Desk* (*CSERD*) (<http://www.shodor.org/refdesk/Resources/Tutorials/PhaseChanges/>). In each DL, researcher-generated homepages were used as conduits to the information provided by the full DLs. The

history homepage consisted of a hyperlink to the Driving Through Time website, as well as six labeled pictures that were hyperlinks to preselected sources within the digital library: photos, newspaper articles, maps, oral histories, drawings, and letters. Similarly, the science DL was organized by resource type: videos, interactive simulations, readings, and virtual tutorials. Participants had access to each DL in its entirety, but most focused on the resources related to the learning task.

3.3 Measures Used in Both Studies

Several instruments were used to gather information about the participants, including a demographic questionnaire, a pretest, and a posttest.

3.3.1 Demographic questionnaire. Each participant was asked to fill out a questionnaire that solicited basic demographic information (e.g., gender, age, academic major).

3.3.2 Pretest and posttest knowledge measures. Paper-and-pencil pretests and posttests were used to capture participants' declarative and conceptual knowledge regarding the domain-specific learning task to which they were randomly assigned. Within each domain, participants completed identical pretest and posttests, containing items designed to measure both declarative and conceptual knowledge. The pretest and posttest in the science domain included ten multiple-choice and six fill-in-the-blank questions. These measures also included an essay in which participants were prompted to interpret a phase change graph and describe what occurs at a molecular level during phase changes (see Appendix C).

The pretest and posttest in the history domain included multiple-choice questions and fill-in-the-blank questions regarding the Blue Ridge Parkway. The measures also included an essay in which participants were prompted to provide information regarding the historical context of the Blue Ridge Parkway, including topics such as groups affected by the Parkway and the

significance of the Parkway to its historical era (see Appendix D). Cronbach's alpha was acceptable for each pretest measure (see Table 1). Posttest alphas were lower, likely due to the lower standard deviation. These lower standard deviations are likely due to participants' knowledge gains while using the DL. Our desire to use the same instruments at pretest and posttest, so that we could statistically analyze learning gains, outweighed concerns about reduced variance, and subsequent effects upon alpha, at posttest.

3.4 Coding and Scoring Used for Both Studies

3.4.1 Knowledge measures. Three members of the research team independently scored the multiple-choice items using a pre-established answer key, and then met to determine whether there were any errors in scoring. The fill-in-the-blank and essay items were scored using criteria and rubrics developed by the researchers regarding the specificity, thoroughness, and quality of the responses. The areas of knowledge measured in the history essay included the people involved in the history of the Blue Ridge Parkway, places included in Parkway, the period of time in which the Parkway was built, the purpose of the Parkway, the positive and negative effects of building the Parkway, and the participants' understanding of the historical context of the Parkway. For the historical context aspect of the essay, the researchers read for three main clusters of ideas regarding the purpose of the Parkway: the idea that it was a scenic highway connecting two national parks, that it was a boom to tourism, and that it helped to preserve the landscape. Participants received between zero to three points depending on how many ideas regarding the purpose of the Parkway they included in their essays. For the science essay, the scoring rubric included four key points related to the apparent paradox of observed constant temperature while objects are being heated at the point of phase change. Participants were assessed on a continuum from no mastery to full mastery based on how clearly they were able to

convey that heat energy was being used to break the bonds of a substance's particles during a phase change, thus resulting in a constant observed temperature. Again, there were three researchers who scored essays. Each essay was scored by two of these researchers, independently. Then, the researchers met to reconcile their scores. If the researchers were unable to resolve their scoring differences, then the first author was consulted to assist in finding an appropriate resolution. The first author was consulted to resolve scoring differences only five times over the course of scoring for both Study 1 and Study 2.

3.4.2 SRL coding. Researchers coded transcriptions of TAPs for SRL processes using a revised coding scheme originally developed by Azevedo and colleagues (Azevedo & Cromley, 2004; Azevedo, Guthrie, & Seibert, 2004). Based upon our review of literature on learning in science and history, and our evaluation of TAPs from the pilot study, the scheme was augmented with science and history-specific strategy codes (e.g., “establishing chronology” and “historical perspective taking” were added to the history DL scheme, and “comparing and contrasting” was added to the science DL scheme). Additionally, several codes were added to both the science and history coding schemes including: corroborating sources, hypothesizing, evaluating content as relevant to task goal, inferring source content, manipulating representation, representation difficulty, and task difficulty. The changes to the codebooks resulted in the history coding scheme having 46 codes and the science coding scheme having 45 codes, each representing distinct micro-level SRL processes. Each of the micro-level codes was further categorized into one of five macro-level processing codes: *planning*, *monitoring*, *strategy use*, *task difficulties and demands*, and *interest* (see Appendix A).

Participants' verbalizations were separated into segments before the coding process. For a segment to be codeable, it had to correspond to one of the micro-level self-regulatory processes

outlined in the codebooks. For example, if a participant said, “I don’t understand the role of latent heat in the phase change process, but I don’t think this video is going to help me with that,” the sentence would be divided into two segments, each of which would be coded as a different SRL process. “I don’t understand the role of latent heat in the phase change process” would be coded as judgment of learning minus (JOL-, i.e., the participant judged their learning to not be adequate regarding a portion of the learning task; see Appendix A). The second segment “but I don’t think this video is going to help me with that” would be coded as expectation of adequacy of content minus (EAC-, i.e., the participant did not feel that the content would be adequate given the goal).

The participants’ TAPs were transcribed using the audio and video recordings. Every utterance spoken by the participants or the researcher was recorded in the transcripts. Italics were used for sections of the transcript in which the participant read directly from a webpage in one of the digital libraries. This was done to more easily identify the participant’s codeable segments. We did not code when participants simply read text from the DL. However, we did code when a participant re-read a section of five or more words, in which case these segments were coded as re-reading (see Appendix A). Some segments did not contain any codeable information (i.e. they did not include information that reflected any of the micro-level self-regulatory processes depicted in the codebooks). These segments were categorized as “not codeable” and were not included in the analysis. Three research team members were trained by the first author to code. Each transcription was independently coded by two team members, who then met to reconcile differences. Also, each transcript was compared to the video recordings to enhance the accuracy of the coding process.

Interrater agreement. As previously noted, each transcription was independently coded twice, by two members of a three-person research team. The two researchers subsequently met to resolve any discrepancies in their coding. The first author was consulted if the two research team members could not reach consensus for a particular code. This type of “two-pass” coding (i.e., two research team members code each transcription independently) is based on the work of Chi (1997) and similar to procedures used by Herrenkohl and Cornelius (2013). These procedures take into account the complexity of coding, on average, 70 verbalized segments per transcription using a coding scheme of 45 or more codes. This procedure places emphasis on ensuring that at least two researchers agreed that a particular segment should be coded in a particular way, and this procedure ensures greater validity than methods that result in a single researcher coding some subset of the transcriptions, as is typical in studies that report interrater reliability. Previous research using this coding scheme, and subsets of the total data, has shown that independent coders can achieve acceptable reliability (e.g., Greene & Azevedo, 2009).

However, as Hammer and Berland (2014) have argued, reliability in coding may not be nearly as important as predictive validity. While raters can be trained to reliably code verbalizations, if those codes have no predictive validity, then the reliability means little. In our case, this coding scheme has been used in numerous studies (e.g., Greene, Costa, et al., 2010; Greene et al., 2012; Greene, Yu, et al., 2014) with results that align well with conceptual models of SRL and hypotheses derived from those models. These past studies, and their findings, support the predictive validity of the coding scheme. In addition, statistical calculations of interrater agreement were not appropriate for this study because two research team members coded every segment of every transcription independently, and then reconciled their coding, with any disagreements resolved through consultation. Reliability measures are more appropriate

when raters code some subset of the data without another person verifying their coding, which was not the case in this study. Calculations of SRL TAP data coding agreements are rarely performed due to the complexity of coding and the privileging of accuracy over interrater reliability. When they are calculated they tend to be conducted on a very small subset of the total corpus of data (Bannert, Reimann, & Sonnenberg, 2014). Our “two-pass” method, with two coders agreeing on every single code assigned to every transcription, reflects our emphasis on validity, with its associated increased resource demands, over interrater reliability.

On the other hand, a related but separate concern is whether other researchers or research teams can reliably implement the coding scheme. Various research teams have used coding schemes derived from Azevedo and Cromley (2004), and the numerous published studies that have resulted (e.g., Greene, Yu, et al., 2014; Johnson, Azevedo, & D’Mello, 2011; Moos, 2013) are one source of evidence that the coding scheme can be reliably implemented by many different groups of people. Likewise, the recent proliferation of research with TAPs that do not derive from the Azevedo and Cromley (2004) work, but with similar coding schemes and interrater procedures (e.g., De Backer, Van Keer, & Valcke, 2014; Dinsmore & Alexander, 2015), is another source of evidence that these procedures can produce inferences with strong validity and reliability evidence.

3.5 Data preparation. Several of the micro-level codes were not used in the analyses including: tutor-initiated time monitoring (TITM), emotion regulation (ER), and help-seeking behavior (HSB). ER and HSB were not used because few to no participants verbalized statements that could be coded as such; TITM was not used because it was irrelevant to participant SRL use.

4. Study 1

4.1 Method

4.1.2 Participants. During the 2012-2013 academic year, 94 students at a university in the southeastern United States received extra credit in an Education course in exchange for their participation in this study. Due to technical issues, data for three participants were lost. Participants' ages ranged from 18 to 23 with a mean age of 20.03 ($SD = .95$). There were 57 females and 34 male students, including 6 freshmen, 16 sophomores, 54 juniors and 15 seniors. Average reported GPA was 3.35 ($SD = .35$) with a range of 2.6 to 3.98. The top three majors were Education (19), Math (10) and Biology (9). The participants were randomly assigned to either the science DL ($n = 47$) or the history DL ($n = 44$). All participants received, read, and signed an informed consent document outlining the study, which was approved by the University's Institutional Review Board.

4.2 Results

4.2.1 Descriptive statistics. Examination of descriptive statistics showed that, on average, participants' mean knowledge scores increased from pretest to posttest (see Table 1). This was true for both the science domain and the history domain. Distributions appeared relatively normal for all measures in both domains. In the science domain, on average, participants enacted macro-level monitoring SRL processes most often, whereas in history the most frequently enacted macro-level SRL processes were strategies.

4.2.2 Learning gains. Research question 1.1 concerned whether participants' knowledge increased from pretest to posttest. On average, participants had higher scores on the posttest than on the pretest in both DLs, and dependent t-tests indicated that these gains were statistically and practically significant (see Table 2). These results support the hypothesis that participants, on average, did learn during their time with the DL.

4.2.3 SRL processing and learning gains. Research questions 1.2 and 1.3 addressed the degree to which participants' macro-level SRL processing predicted post-test knowledge scores. Given our interest in the domain-specific aspects of SRL, we explored two different techniques for computing macro-level SRL variables (Greene et al., 2013; Greene, Yu, et al., 2014). The first technique mirrored the work of Greene and Azevedo (2009) and involved aggregating all of the micro-level SRL processes into their respective macro-level SRL variables (i.e., full aggregation). The second technique involved exploratory analyses to determine which micro-level SRL processes were most related to posttest performance, and then aggregating just those processes (i.e., data-driven aggregation). We compared the predictive validity of these techniques as a way of providing initial empirical evidence regarding (a) the question of domain-specificity of SRL, as well as (b) the optimal way to model SRL in our quantitative analyses.

4.2.3.1 Full aggregation technique. Using this technique, all micro-level SRL processes were aggregated into three macro-level SRL variables (see Appendix A) using simple sums per the work by Greene and Azevedo (2009). This resulted in macro-level SRL variables being created for planning, monitoring and strategy use. Descriptive information for these variables is presented in Table 1¹. To test research question 1.2, we regressed posttest scores on their respective pretest scores and the three macro-level SRL variables, by DL (see Table 3). Results of these analyses showed that full aggregates of the micro-level SRL codes into macro-level SRL variables were not statistically significant predictors of posttest scores above and beyond pretest, and indicated no support for the hypothesis that the full-aggregation macro-level SRL variables were useful predictors of learning.

---

¹ Correlation matrices are available upon request to the first author.

4.2.3.2 Data-driven aggregation technique. Using correlation analysis¹, we identified micro-level SRL processes that were positively related to knowledge scores in each DL, as well as those that were negatively related. Then we aggregated only those micro-level SRL processes with an absolute value of r greater than or equal to .1 (see Table 4). For example, in the science DL, plan and judgment of learning plus, among other micro-level SRL variables, were found to have a positive correlation with posttest knowledge scores equal to or greater than .1. Therefore the counts for these variables, among others with similar correlations, were summed to create the “SRL positive” macro-level variable for science. On the other hand, in the science DL, judgment of learning negative and rereading, among other micro-level SRL codes, were negatively correlated with posttest knowledge scores at a value of negative .1 or less (i.e., from negative .1 to negative 1), and therefore participants’ counts on these variables were all summed into a “SRL negative” variable for the science DL. We aggregated unique macro-level SRL positive and negative variables for each DL (see Table 1 for descriptive information on these variables). Posttest scores were regressed onto both the pretest scores and the new data-driven aggregation macro-level SRL variables (e.g., SRL positive and SRL negative), with separate analyses for each DL.

The final sets of micro-level SRL codes comprising each macro-level SRL variable are shown in Table 4, and the results of the analyses with the SRL positive and negative variables are reported in Table 5. All but one of the positive and negative data-driven aggregation macro-level SRL variables were statistically significant at the $p < .05$ level, with relations in the expected direction (e.g., the negative SRL variables had negative beta weights). In addition, r -squared values were higher for models including variables derived from the data-driven aggregation methods, compared to the full aggregation methods. These results show that the

data-driven aggregation method of macro-level SRL creation produced a model that was a better fit for our data both in terms of statistical significance (i.e., based on p -values) and practical significance (i.e., based on beta weights and effect sizes), compared to full aggregation.

Some common findings across DLs were evident (see Table 4). More frequent verbalizations of corroborating sources, positive feelings of knowing, knowledge elaboration, memorization, prior knowledge activation, self-knowledge activation, and time planning were all associated with learning gains in both DLs. Many of these micro-level SRL processes are considered high-level strategies (Dunlosky et al., 2013; Schraw, 2006) that are logical predictors of positive learning gains regardless of academic domain. Notably, in general, micro-level SRL processes within the planning and monitoring macro-level SRL categories did not consistently predict learning gains across DLs. Another finding was that some micro-level processes that were grouped together in the original Greene and Avezedo (2009) macro levels were found to be positive predictors in some cases and negative predictors in others. For example, making a plan and evaluating content as relevant to the task were found to be positive predictors of knowledge gains in science but negative predictors in history. These results indicate that some of the processes that were included in the full aggregation method, all of which were assumed to be positive predictors of learning, may actually not have been positive, and therefore reduced the predictive power of the macro-level SRL variables in which they were aggregated.

4.2.3.3 Summary of comparison of full and data-driven aggregation techniques. In summary, to address research questions 1.2 and 1.3, the best method of aggregating the micro-level SRL processes for analysis was the data-driven method. The full aggregation method of computing macro-level SRL variables, used by Greene and Avezedo (2009), did not seem to be sufficiently nuanced for this sample or the multiple contexts. Across both DLs and knowledge

measures within DLs, there were differences in the specific micro-level SRL processes that proved predictive of learning gains, but these differences were greater among planning and monitoring processes, compared to strategy use.

4.2.4 Comparison of self-regulated learning processing across domains. The final research question for this Study, 1.4, concerned the degree to which the frequency of micro-level SRL processing, on average, differed across domains. Given that we randomly assigned participants to DLs, differences in average counts of SRL processing can be inferred as evidence of differences in SRL processing across domains. Table 6 shows descriptive statistics for the frequency of each micro-level SRL process, across DLs. To investigate this hypothesis, we followed the procedures outlined by Greene, Costa, and Dellinger (2011). First, we determined which count distribution (i.e., Poisson, Negative-Binomial, or zero-inflated versions of those distributions) best fit each micro-level SRL variable using AIC, BIC and Sample Size Adjusted BIC values². Once the best fit of these four distributions was found, count regressions were run where individual micro-level SRL variables were regressed onto a categorical DL variable, with history as the comparison group. Items that were unique to a domain (e.g., compare and contrast, judgment of learning tentative, historical perspective seeking and establishing chronology, see Table 6) were excluded from these analyses. Given the many analyses conducted, the Benjamini and Hochberg (1995) method of controlling for multiple comparisons, by using the false discovery rate, was used. After the corrections, there were 12 statistically significant differences at the $p < .05$ level (see Table 7). On average, participants assigned to the science DL more frequently enacted several different types of monitoring micro-level SRL processes (e.g., content evaluation, judgment of learning plus), compared to participants assigned to the history DL. On

---

² Comparisons of count distribution models for each micro-level SRL variable available upon request to the first author

the other hand, participants in the history DL, on average, more frequently enacted knowledge elaborations, took notes, and expressed interest in the task or DL.

4.3 Discussion of Study 1 Findings

Our results provided evidence that participants did benefit from utilizing both DLs, as indicated by statistically and practically significant knowledge gains. Our findings support the work of Alexander et al. (2011) and Greene et al. (2013), that, indeed, investigations of the domain-specificity of SRL are needed, as the full aggregation method of calculating macro-level SRL variables was not sufficiently nuanced for these domains, tasks, and learning environments. Instead, data-driven aggregation methods were the best fit to the data. Domain-differences were further evidenced in the analyses comparing average frequency of use of various micro-level SRL processes across DLs. Again, given random assignment of participants to DLs, any statistically significant differences in average count of SRL processing, controlling for Type I error, are suggestive of substantive differences in how SRL occurs across domains. On average, the science DL elicited greater monitoring, particularly of the content, than the history DL. This finding, convergent with the results of the previous analyses, further supports the argument that aspects of SRL processing may be domain-specific, if not task and environment specific as well.

The argument for using domain-specific, data-driven aggregation methods to model SRL processing is to some degree undermined by the post hoc nature of the data-driven aggregation method. To address this concern, we decided to attempt to partially replicate the data-driven aggregation method macro-level SRL variables created in Study 1, using data from our pilot, which for clarity in presentation here we are calling Study 2.

5. Study 2

Our second study was a partial replication of Study 1. In this Study 2, we aggregated the same micro-level SRL processes from Study 1 into the same macro-level SRL variables, using a different dataset (i.e., we used the same variables as listed in Table 4). Then we conducted the same analyses as in Study 1, regressing posttest scores on pretest scores and our macro-level SRL process variables. This partial replication allowed us to investigate whether our data-driven aggregation method and findings from Study 1 would replicate with a new sample, and if so, such results would bolster the argument that the findings were substantive as opposed to being the result of sample-specific random covariances in the data. Replications are an important step in normative social science, particularly when new methods are developed and tested (Makel & Plucker, 2014).

5.1 Method

5.1.1 Participants. During the 2011-2012 academic year, 30 students at a university in the southeastern United States received extra credit in an Education course in exchange for their participation in this study. Participants' ages ranged from 18 to 23 with a mean age of 19.90 ($SD = 1.21$). There were 18 females and 12 male students, including 12 freshmen, 8 sophomores, 7 juniors and 3 seniors. Average reported GPA was 3.19 ($SD = .53$) with a range of 2.30 to 4.00. The top three majors were Political Science (4), Psychology (3) and Mathematics (3). The participants were randomly assigned to either the science DL ($n = 15$) or the history DL ($n = 15$). All participants received, read, and signed an informed consent document outlining the study, which was approved by the University's Institutional Review Board.

5.1.2 Measures. The only difference in measures across studies was that in Study 1 the history knowledge measure had eight multiple-choice and two fill-in-the-blank questions, compared to ten multiple-choice questions only in Study 2.

5.1.3 SRL Coding. A small number of micro-level SRL codes were added between Study 2 and Study 1. Some of these codes were included in the data-driven aggregation method variables in Study 1. Therefore, we simply omitted these codes in this Study 2 analysis; see Table 4 for more information.

5.2 Results

5.2.1 Descriptive statistics. Examination of descriptive statistics showed that, on average, participants' mean scores increased from pretest to posttest for both DLs (see Table 8). Distributions appeared relatively normal for all knowledge measures.

5.2.2 Learning gains. Research question 2.1 concerned whether participants' time with the DL was associated with gains in knowledge. Dependent t-tests indicated that, for both DLs, knowledge gains were statistically and practically significant (see Table 9). These results indicate that participants, on average, did learn during their time with DL.

5.2.3 SRL processing and learning gains. Research question 2.2 concerned whether the SRL positive and negative variables, created using the data-driven aggregation methods used in Study 1, would replicate with a new sample. We calculated positive and negative macro-level SRL variables for each DL, using the same micro-level SRL processes as in Study 1 (see Table 4 for exceptions). Posttest scores were regressed on pretest scores, and the positive and negative SRL variables (see Table 10). Given the small sample size and low power in this study, we did not expect to find statistically significant effects. Indeed, even the pretest variable was not a statistically significant predictor of posttest score in the model with the data-driven positive and negative macro-level SRL variables. On the other hand, perhaps most importantly, the beta weights for each of our SRL variables were practically significant, with all regression coefficients in the expected direction.

5.3 Discussion of Study 2 Findings

Overall, despite relatively low power, Study 2 proved to be an effective sample for partial replication purposes. The data-driven macro-level SRL variables computed in Study 1 displayed similar predictive validity in Study 2, with a different sample. These findings provide further evidence to bolster the argument that the data-driven aggregation method produced macro-level SRL variables that were helpful predictors of learning, and not solely due to random chance. Further, this partial replication, when considered in conjunction with Study 1, demonstrates how the data-driven aggregation method can be used to investigate issues of domain-specificity across DLs and domains.

6. General Discussion

The affordances of 21st century digital learning environments bring with them unique demands upon learners' ability to navigate vast and complex knowledge representations, as well as their ability to vet and integrate those representations into deep conceptual understanding of content in academic domains (Azevedo, 2005; Collins & Halverson, 2009; Greene, Moos, et al., 2011; Strømsø & Bråten, 2010). Unfortunately, many learners struggle to effectively learn with digital environments like the Internet (Bennett et al., 2008). Past empirical research has indicated that learners' success is likely dependent on how well they can self-regulate their cognition, metacognition, motivation, affect, and behaviors (Azevedo, Johnson, et al., 2010; Greene, Bolick, et al., 2010; Greene, Costa, et al., 2010). Yet, if the ultimate purpose of this line of research is to develop methods of fostering SRL among those learners who do not yet effectively enact it on their own, at some point the question of domain-specificity must be addressed (Alexander et al., 2011). Specifically, what aspects of SRL, if any, are effective in multiple domains, and which are truly unique to tasks in science, history, or other academic disciplines

(Greene et al., 2013)? Investigations into this question are made all the more challenging by the necessity of using concurrent online SRL measures (Winne & Perry, 2000), such as TAPs, and the resource demands they present (Greene, Robertson, et al., 2011). Therefore, this research study had two foci: theoretical questions about the domain-specificity of SRL, and methodological questions about how best to capture and model SRL TAP data to address the first focus. In two studies, we randomly assigned college students to one of two DLs, and used TAPs to capture how they self-regulated while learning about topics in science or history. With these data, we were able to investigate how SRL processing differed within and across domains, providing initial evidence regarding the degree to which researchers and educators should be conceptualizing self-regulation as domain-specific.

Our findings indicated that participants showed learning gains in both science and history. However, simply aggregating every micro-level SRL process into macro-level SRL variables (i.e., full aggregation) did not prove to be a useful way of predicting performance. Rather, a data-driven aggregation approach best fit the data and revealed that the SRL processes associated with learning gains differed across domains. Therefore, these findings provide initial support for Greene and colleagues' (2013) claim that understanding which SRL processes to foster critically depends upon the domain in which learners are working, and perhaps even the task they have been assigned.

More specifically, across both of the domains and tasks assigned in this study, there were some strategies whose frequency of use predicted learning gains. These "high-level" strategies (Schraw, 2006) included corroborating sources, knowledge elaborations, and activating prior and self-knowledge. However, in line with Alexander and colleagues' (2011) arguments, other high-level strategies were domain-specific (e.g., historical perspective taking and inferring source

content in history). Further, there were a number of monitoring and planning processes that differed in their predictive validity across the two CBLEs. In science, evaluating the relevance of content and making plans positively predicted learning gains, whereas the frequency of this process negatively predicted learning in history. On the other hand, monitoring the use of strategies and self-questioning were positive predictors of learning in history, but negative predictors in science. Interestingly, judgments of learning were highly predictive of learning gains in science, but not so in history. These findings indicate that more research is needed into the situated aspects of SRL processing (Sandoval, 2012).

It was also notable that several strategies were negatively correlated with learning, including rereading, and reading notes, and frequently selecting new information sources (see Table 4). These findings align with scholarship regarding low-level strategies, and their relative lack of efficacy (e.g., Dunlosky et al., 2013; Schraw, 2006). Clearly, it is not the case that every strategy is a helpful strategy, and our work takes a first step toward identifying those strategies that might have similar, or differential, utility across domains.

What is to be made of the finding that certain SRL processes differed in predictive validity, and even utility (i.e., positive versus negative correlations with learning gains) across domains? We suggest caution before extrapolating too much from the findings of one study involving two domains. But, we do believe these findings provide support for the assertions of Alexander and colleagues (2011), among others, that issues of domain-specificity in SRL indeed do matter and should be investigated. Learners have limited time and resources, and it is likely the behaviors that lead to comprehension of one phenomenon may differ from those best suited to understanding a different phenomenon. It is our hope that future research will reveal SRL processes with broad utility (e.g., knowledge elaboration) that can be taught as heuristics, as well

as those SRL processes that might be unique to particular contexts, but no less important within those contexts. The utility of various SRL processes likely also varies in terms of learners' prior knowledge (Moos & Azevedo, 2008).

Finally, the random assignment of participants to CBLEs allowed for a direct comparison of the frequency of particular micro-level SRL process use across DLs, and we found evidence that participants more often monitored their learning in science, and more often engaged in knowledge elaboration in history. While we believe that our findings make a substantive contribution to the literature on SRL and domain-specificity, and point to several seemingly productive lines of future research, we also acknowledge that the exploratory nature of this work precludes any grand generalizations about optimal SRL processing, within or across academic domains.

6.1 Limitations

Our studies were some of the first to use online measures to examine the degree of domain-specificity of SRL, and therefore like any initial research, ours has particular limitations. While across our studies we gathered a particularly large set of TAP data, we still encountered power issues. Ideally, our replication sample (i.e., Study 2) would have had a sample size sufficient to provide both practically and statistically significant results. Therefore, it is important to note that our replication study provides only limited evidence, and that any external validity claims would require additional research with larger samples. Gathering, processing, and analyzing large samples of TAP data comes with enormous resource demands, therefore multiple teams of researchers should consider consolidating or combining their efforts to rigorously design and gather large datasets that can be used to address questions of replicability of findings regarding efficacious SRL processing. The dissemination of our findings, even with the

limitations associated with replication study, extends current literature and will allow other researchers to build off of this work to investigate ways of maximizing power and resource cost (Makel & Plucker, 2014).

There are benefits to randomly assigning participants to CBLEs and domains in terms of making inferences about the frequency of micro-level SRL process enactment. On the other hand, the lack of a within-subjects design did not allow for an investigation of how participants themselves varied their SRL across domains. Future research should assign participants to different domains and/or different tasks within a domain, and use online measures to examine intraindividual variation, or lack thereof, in SRL processing. Further, a between-subjects design also precluded an examination of how prior knowledge might relate to differences in SRL processing across domains (see section 4.2.4). Clearly, the substantial evidence of the prominent role of prior knowledge in SRL processing (e.g., Moos & Azevedo, 2008) warrants continued investigation of how it might relate to SRL processing differences across domains, using within-subjects designs that allow for an examination of how participants' prior knowledge, in each domain, predicts SRL in those domains.

While we were able to vary the domains in this study, we did not vary the tasks themselves. As Greene and colleagues (2013) have stated, aspects of SRL may be not only domain-specific, but task-specific as well. A study that incorporates the assignment of participants to multiple tasks within the same domain would allow researchers to investigate the degree to which SRL is not only domain-specific, but also even situated and contextual.

The SRL TAP coding scheme developed by Azevedo, Greene, and colleagues (see Azevedo, 2005 and Greene, Robertson, et al., 2011 for details) has proven to be a powerful tool. However, online, concurrent measures of SRL do require asking participants to engage in

verbalizations while the relevant cognition occurs, and may therefore have an effect upon processing that retrospective measures, such as surveys, do not. There is evidence that asking participants to verbalize their thinking, rather than explain it, does not have appreciable effects upon cognition (cf. Fox, Ericsson, & Best, 2011; Greene, Robertson, et al., 2011; Veenman, 2005; Veenman et al., 1993), but we must acknowledge that in every study using TAPs, there is the threat of such effects. Researchers should be vigilant when using TAPs, to ensure that they are structured in ways that minimize the chance of effects upon cognition (Fox, 2009), and they should consider triangulating TAPs with other offline or retrospective measures, as a means of establishing concurrent validity evidence. The reliability and replicability of TAP coding are also concerns (Greene, Robertson, et al., 2011). More research, involving multiple studies using the same TAP coding scheme, as has been done with the Azevedo scheme (e.g., Azevedo, Moos, et al., 2010), is necessary to demonstrate the affordances and limits of this method of capturing SRL (Winne & Perry, 2000).

Finally, TAPs are one of many possible ways of capturing SRL through online measures. Researchers should consider a wider array of online measures of SRL, including trace data, prompt data, and the construction of environments where learner choices indicate SRL processing (Greene, Muis, & Pieschl, 2010). Again, the exploration of the domain-specificity of SRL will require numerous studies and samples, and the diversification of the measurement techniques used in those studies will allow for more defensible inferences about the phenomena under study.

6.2 Future Directions

Aside from the recommendations for future research stated previously, the field of SRL would benefit from more nuanced interventions designed to foster SRL (Dignath & Büttner,

2008). As our data suggest, SRL interventions that simply advise students to “monitor” may not provide a sufficient level of detail, or scaffolding, to help the students enact that advice. Direct instruction and support on what to monitor (i.e., resource quality, growing understanding) and how to monitor (i.e., how to automate metacognitive monitoring) may be necessary. Likewise, our data suggest that high-level strategies may vary by domain, therefore what is taught as an elaborative and constructive strategy in history (e.g., historical perspective taking) may by necessity be very different than what should be taught in science. SRL interventions have largely focused on the various phases of SRL (e.g., Winne & Hadwin, 2008; Zimmerman, 2000) but students may need direct instruction on how to specifically plan, monitor, and use strategies for a particular task, in a particular domain.

However, despite our findings that there are domain-specific aspects of SRL, researchers and educators would be well-served by additional exploration of whether, and how, SRL transfers across domains and tasks. It seems intuitive that some aspects of self-regulatory processing should prove useful across domains, but the key is to determine which aspects, and how to help students realize that such transfer is beneficial to them. Future research should investigate how instruction on SRL in a particular domain or task influences SRL processing in other domains and tasks.

6.3 Conclusion

We believe that the studies outlined here represent an important first step toward extending the literature by addressing Alexander and colleagues’ (2011) call for research into the domain-specific aspects of SRL. Research in this area is needed to determine what to include in SRL interventions, and what aspects of SRL are likely to be effective across domains, tasks, and contexts. Our results indicate differences in the predictive validity of particular monitoring

processes in science, as well as differences in the frequency of use of high-level strategies in both science and history. Our replication study provides initial support for our data-driven aggregation technique as a way of identifying key SRL processes for particular domains and tasks. Overall, our studies suggest that there is great promise in interventions that specifically instruct students on planning, monitoring, and high-level strategy use, when tailored to particular domains and tasks.

Authors' note: blinded for review; Editors, please see title page authors' note

References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183-198.
- Alexander, P. A. (1995). Superimposing a situation-specific and domain-specific perspective on an account of Self-Regulated Learning. *Educational Psychologist, 30*(4), 189-193.
- Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2011). Self-regulated learning in academic domains. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance*. New York: Routledge.
- Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist, 40*, 193-197.
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523-535.
- Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and learning, 2*(2-3), 67-87.
- Azevedo, R., Guthrie, J.T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research, 30*(1), 87-111.
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. In *New Science of Learning* (pp. 225-247). New York: Springer.
- Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. C. (2008). Why is externally regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*(1), 45-72.
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*(4), 210-223.
- Baddeley A. D. (2001) Is working memory still working? *American Psychologist, 56* (11), 851-64.
- Bannert, M. & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3*(1), 39-58.

- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition & Learning*, 9, 161-185.
- Bartels, J. M., & Magun-Jackson, S. (2009). Approach–avoidance motivation and metacognitive self-regulation: The role of need for achievement and fear of failure. *Learning and Individual Differences*, 19(4), 459-463.
- Bembenutty, H., Cleary, T. J., & Kitsantas, A. (Eds.) (2013). *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman*. Charlotte, NC: Information Age Publishing.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57(1), 289–300. doi:10.2307/2346101
- Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775-786.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, 31(6), 445-457.
- Bolick, C. M. (2006). Digital archives: Democratizing the doing of history. *The International Journal of Social Education*. 21 (1), 122-135.
- Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R. & Mautone, P. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria Digital Earth Prototype (ADEPT). *Library Trends*, 49(2), 228-250.
- Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist*, 49(2), 59-69.
- Burton, O. V. (2005). American digital history. *Social Science Computer Review*, 23(2), 206-220.
- Carretero, M., & Lee, P. (2014). Learning historical concepts. In R. K. Sawyer (Ed.), *The Cambridge Handbook of The Learning Sciences* (2nd Ed.) (pp. 587-604). New York: Cambridge University Press.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, 6, 271–315.
- Chi, M.T.H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of Learning Sciences*, 14(2), 161-199.

- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Cognition and Instruction*, 19, 323-393.
- Collins, A. & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. New York: Teachers College Press.
- De Backer, L., Van Keer, H., & Valcke, M. (2014). Promoting university students' metacognitive regulation through peer learning: the potential of reciprocal peer tutoring. *Higher Education*, 1-18. doi:10.1007/s10734-014-9849-3
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning*, 3, 231-264.
- Dinsmore, D. L., & Alexander, P. A. (2015). A multidimensional investigation of deep-level and surface-level processing. *Manuscript accepted for publication in The Journal of Experimental Education*.
- Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40, 117-128.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T., (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. (2007). *Taking science to school: Learning and teaching in grades K-8*. Washington, D.C.: The National Academies Press.
- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223-242). Cambridge, MA: Cambridge University Press.
- Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data (2nd ed.)*. Cambridge, MA: MIT Press.
- Fox, E. (2009). The role of reader characteristics in processing and learning from informational text. *Review of Educational Research*, 79(1), 197-261.
- Fox, M.C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137, 316-344.

- Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: Supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development*, 56(1), 73-92.
- Greene, J. A., & Azevedo, R. (2007). Adolescents' use of self-regulatory processes and their relation to qualitative mental model shifts while using hypermedia. *Journal of Educational Computing Research*, 36(2), 125-148.
- Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of sophisticated mental models. *Contemporary Educational Psychology*, 34, 18-29.
- Greene, J. A., Bolick, C. M., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: The role of self-regulated learning. *Computers & Education*, 54, 230-243.
<http://dx.doi.org/10.1016/j.compedu.2009.08.006>
- Greene, J. A., Costa, L.-J., & Dellinger, K. (2011). Analysis of self-regulated learning processing using statistical models for count data. *Metacognition & Learning*, 6, 275-301.
<http://dx.doi.org/10.1007/s11409-011-9078-4>
- Greene, J. A., Costa, L. C., Robertson, J., Pan, Y., & Deekens, V. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education*, 55, 1027-1043.
[doi:10.1016/j.compedu.2010.04.013](https://doi.org/10.1016/j.compedu.2010.04.013)
- Greene, J. A., Dellinger, K. R., Tüysüzoğlu, B. B., Costa, L. J. (2013) A two-tiered approach to analyzing self-regulated learning data to inform the design of hypermedia learning environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies*. (pp. 117-128). New York, NY: Springer.
[doi:10.1007/978-1-4419-5546-3_8](https://doi.org/10.1007/978-1-4419-5546-3_8)
- Greene, J. A., Hutchison, L. A., Costa, L., & Crompton, H. (2012). Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic. *Contemporary Educational Psychology*, 37, 307-230. <http://dx.doi.org/10.1016/j.cedpsych.2012.02.002>
- Greene, J. A., Moos, D. C., & Azevedo, R. (2011). Self-regulation of learning with computer-based learning environments. *New Directions for Teaching and Learning*, 126, 107-115.
- Greene, J. A., Muis, K. R., & Pieschl, S. (2010). The role of epistemic beliefs in students' self-regulated learning with computer-based learning environments: Conceptual and methodological issues. *Educational Psychologist*, 45(4), 245-257.

- Greene, J. A., Robertson, J., & Costa, L-J. C. (2011). Assessing self-regulated learning using think-aloud protocol methods. In B. J. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 313-328). New York: Routledge Publishers.
- Greene, J. A., & Yu, S. B. (2014). Modeling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology*, 39, 12-28.
<http://dx.doi.org/10.1016/j.cedpsych.2013.10.002>
- Greene, J. A., Yu, S., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education*, 76, 55-69.
<http://dx.doi.org/10.1016/j.compedu.2014.03.008>
- Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, 23(1), 37-46.
- Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *The Journal of the Learning Sciences*, 22(3), 413-461.
- Hicks, D., & Doolittle, P. E. (2009). Multimedia-based historical inquiry strategy instruction: Do size and form really matter? In J. Lee & A. Friedman (Eds.), *Research on technology in social studies education* (pp. 127-154). Charlotte, NC: Information Age.
- Hmelo-Silver, C., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *Journal of the Learning Sciences*, 15(1), 53-61.
- Jacobson, M., & Archodidou, A. (2000). The design of hypermedia tools for learning: Fostering conceptual change and transfer of complex scientific knowledge. *Journal of the Learning Sciences*, 9(2), 145-199.
- Jacobson, M., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences*, 15(1), 11-34.
- Johnson, A., Azevedo, R., & D'Mello, S. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, 29(4), 471-504.
- King, R. B., & McInerney, D. M., (2014). Culture's consequences on student motivation: Capturing cross-cultural universality and variability through personal investment theory. *Educational Psychologist*, 49(3), 175-198, doi:10.1080/00461520.2014.926813
- Koedinger, K., Aleven, V., Roll, I., Baker, R. (2009). In vivo experiments on whether supporting

- metacognition in intelligent tutoring systems yields robust learning. In A. Graesser, & D. Hacker (Eds.), *Handbook of Metacognition in Education*. (pp. 383-412). New York, NY: Routledge.
- Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, 1-15.
- Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology (2nd ed.)* (pp. 803-821). Mahwah, NJ: Erlbaum.
- Lee, J. K., (2002). Digital history in the history/social studies classroom. *History Teacher*, 35(4), 503-517.
- Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, 46(9), 1023-1040.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43, 304-316.
- Mason, C., Berson, M. J., Diem, R., Hicks, D., Lee, J. K., & Dralle, T. (2000). Guidelines for using technology to prepare social studies teachers. *Contemporary Issues in Technology and Teacher Education*, 1(1), 107-116.
- Mayer, R. (2005). *The Cambridge handbook of multimedia learning*. NY: Cambridge University Press.
- Mervis, J. (2009). Study questions value of school software for students. *Science*, 323(5919), 1277. doi:10.1126/science.323.5919.1277
- Moos, D. C., (2013). Examining hypermedia learning: The role of cognitive load and self-regulated learning. *Journal of Educational Multimedia and Hypermedia*, 22(1), 39-61.
- Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, 33, 270-298.
- Nasah, A., DaCosta, B., Kinsell, C., & Seok, S. (2010). The digital literacy debate: An investigation of digital propensity and information and communication technology. *Educational Technology Research and Development*, 58(5), 531-555. doi:10.1007/s11423-010-9151-8.
- National Research Council (2005). *How students learn: Science in the classroom*. Washington, DC: The National Academies Press.

- Osborne, J. F., Ratcliffe, M., Collins, S., Millar, R., & Duschl, R. (2003). What 'ideas-about-science' should be taught in school science? A delphi study of the 'Expert' community. *Journal of Research in Science Teaching*, 40(7), 692–720.
- Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York: Basic Books.
- Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: the cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning*, 8(3), 213-234.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Prensky, M. (2001) Digital natives, digital immigrants. *On the Horizon*, 9, 1-6.
- Rotgans, J. I., & Schmidt, H. G. (2009). Examination of the context-specific nature of self-regulated learning. *Educational Studies*, 35(3), 239-253.
doi:10.1080/03055690802648051
- Sandoval, W. A. (July, 2012). Situating epistemological development. In J. van Aalst, K. Thompson, M.J. Jacobson, & P. Reimann (Eds.), *Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) - Volume 1, Full Papers* (pp. 203-209). International Society of the Learning Sciences: Sydney, NSW, Australia.
- Schraw, G. (2006). Knowledge: Structures and processes. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed.) (pp. 245-264). San Diego: Academic Press.
- Selwyn, N. (March, 2009). *The digital native-myth and reality*. Paper presented at 10th Annual CLIP (Chartered Institute of Library and Information Professionals) London seminar series. <http://www.scribd.com/doc/9775892/Digital-Native>
- Strømsø, H. I. & Bråten, I., (2010). The role of personal epistemology in the self-regulation of internet-based learning. *Metacognition & Learning*, 5, 91-111.
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147-177.

- Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und metakognition: Implikationen für forschung und praxis* (pp. 77–99). Münster: Waxmann.
- Veenman, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, 2, 177-183.
- Veenman, M. V. J., Elshout, J. J., & Groen, M. G. M. (1993). Thinking aloud: does it affect regulatory processes in learning? *Tijdschrift voor Onderwijsresearch*, 18, 322-330.
- Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia, PA: Temple University Press.
- Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Erlbaum.
- Winne, P. H., & Jamieson-Noel, D. L. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.
- Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Table 1

Study 1 Descriptive Statistics for Pretests and Posttests

	Mean(<i>SD</i>)	Skewness(<i>SE</i>)	Kurtosis(<i>SE</i>)	Cronbach's alpha
<i>Science</i>				
Pretest	9.77(4.54)	.05(.35)	-1.37(.68)	.73
Posttest	14.89(3.29)	-.75(.35)	.33(.68)	.56
Planning Macro-level SRL	10.78(7.60)	.58(.35)	-.02(.68)	
Strategy Use Macro-level SRL	29.91(13.87)	.52 (.35)	-.33(.68)	
Monitoring Macro-level SRL	34.17(22.91)	1.19(.35)	1.68 (.68)	
Positive SRL	49.70(23.68)	.90(.35)	.25(.68)	
Negative SRL	6.74(5.45)	1.06(.35)	1.27(.68)	
<i>History</i>				
Pretest	8.16(5.65)	.27(.36)	-1.04(.70)	.86
Posttest	20.07(3.23)	-.85(.36)	-.48(.70)	.67
Planning Macro-level SRL	12.41(8.46)	1.03(.36)	.95(.70)	
Strategy Use Macro-level SRL	39.07(18.75)	.20(.36)	-.44(.70)	
Monitoring Macro-level SRL	24.47(15.51)	.63(.36)	-.33(.70)	
Positive SRL	39.61(24.02)	.88(.36)	.82(.70)	
Negative SRL	42.25(17.12)	1.99(.36)	8.65(.70)	

Table 2

Study 1 Dependent T-Tests between Pretests and Posttests

Digital Library	Mean difference	<i>df</i>	<i>t</i>	Cohen's <i>d</i>
Science	5.13	46	9.839*	.68
History	11.91	43	17.08*	.87

*p < .001

Table 3

*Study 1 Full Aggregation Multiple Linear Regression Results, Standardized Regression**Coefficients*

	Science	History
<i>Predictor</i>		
Pretest	.66*	.52*
Planning	.09	-.13
Strategy Use	-.08	.24
Monitoring	.09	-.05
<i>F</i>	7.40	6.33
<i>df</i>	4, 42	4, 39
Total R^2	.41	.39
Adjusted R^2	.36	.33
* $p < .001$		

Table 4

Positive and Negative Macro-level SRL Variables, by Digital Library

Science	History
<i>Positive</i>	<i>Positive</i>
Content Evaluation – (.13)	Content Evaluation + (.39)
Corroborating Sources (.28) ^a	Corroborating Sources (.23) ^a
Evaluating Content as Relevant to Task Goal + (.17) ^a	Feeling of Knowing + (.11)
Evaluating Content as Relevant to Task Goal – (.44) ^a	Feeling of Knowing – (.13)
Emotion Monitoring (.11)	Historical Perspective Taking (.26) ^a
Feeling of Knowing + (.25)	Inferencing (.17)
Judgment of Learning + (.13)	Inferring Source Content (.20) ^a
Judgment of Learning Tentative (.16) ^a	Interest + (.33)
Knowledge Elaboration (.19)	Interest – (.19)
Memorization (.11)	Knowledge Elaboration (.31)
Prior Knowledge Activation (.15)	Memorization (.11)
Plan (.11)	Monitoring Use of Strategies (.27)
Search (.28)	Prior Knowledge Activation (.28)
Selecting New Information Source (.22)	Self-Knowledge Activation (.19)
Self-Knowledge Activation (.21)	Self-Questioning (.31)
Time Planning (.16)	Taking Notes (.20)
	Time Planning (.17)
<i>Negative</i>	<i>Negative</i>
Inferring Source Content (-.12) ^a	Evaluating Content as Relevant to Task Goal + (-.24) ^a
Judgment of Learning – (-.24)	Evaluating Content as Relevant to Task Goal – (-.15) ^a
Monitoring Use of Strategies (-.11)	Emotion Monitoring (-.22)
Re-reading (-.24)	Manipulating Representation (-.23) ^a
Self-Questioning (-.10)	Plan (-.15)
	Recycling Goal in Working Memory (-.24)
	Reading Notes (-.16)
	Selecting New Information Source (-.17)
	Time Monitoring (-.12)

Note: Correlation with posttest in parentheses

^a Not coded in Study 2 so not included in macro-level SRL variables for those analyses.

Table 5

Study 1 Data-Driven Aggregation Multiple Linear Regression Results, Standardized Regression Coefficients

	Science	History
<i>Predictor</i>		
Positive SRL	.21	.44**
Negative SRL	-.35**	-.34**
Pretest	.50***	.39**
<i>F</i>	14.04	14.79
<i>df</i>	3, 43	3, 40
Total R^2	.50	.53
Adjusted R^2	.46	.49
* $p < .05$, ** $p < .01$, *** $p < .001$		

Table 6

Study 1 Micro-Level SRL Process Descriptive Statistics by Digital Library

SRL Process	History		Science	
	Mean	SD	Mean	SD
Compare and Contrast (CC) ¹	-	-	.94	1.45
Content Evaluation Plus (CE+)	.86	1.25	1.72	2.16
Content Evaluation Minus (CE-)	1.91	1.90	3.89	3.77
Coordinating Information Sources (COIS)	.34	.68	.28	.71
Corroborating Sources (COS)	.39	.87	.30	.69
Draw (DRAW)	.00	.00	.15	.47
Expectation of Adequacy of Content Plus (EAC+)	1.11	1.54	1.83	2.55
Expectation of Adequacy of Content Minus (EAC-)	.80	1.05	1.72	2.58
Establishing Chronology (EC) ²	1.70	1.84	-	-
Evaluating Content Related to Task Plus (ECRT+)	1.75	2.67	1.68	1.95
Evaluating Content Related to Task Minus (ECRT-)	1.09	1.79	2.57	2.78
Emotional Monitoring (EM)	.09	.36	.17	.48
Emotional Regulation (ER)	.00	.00	.00	.00
Feeling of Knowledge Plus (FOK+)	4.50	3.71	6.94	6.02
Feeling of Knowledge Minus (FOK-)	2.89	3.76	3.06	2.73
Historic Perspective Taking ²	.32	.74	-	-
Help Seeking Behavior (HSB)	.09	.29	.06	.25
Hypothesizing (HYP)	1.27	1.62	1.51	1.99
Inference (INF)	.84	1.29	.85	1.55

Inferring Source Content(ISC)	1.07	1.28	.70	1.12
Interest Plus (INT+)	4.32	4.93	1.89	3.00
Interest Minus (INT-)	.45	.95	.28	.77
Judgment of Learning Plus (JOL+)	.64	1.10	2.68	3.81
Judgment of Learning Minus (JOL-)	1.77	1.71	2.60	3.29
Judgment of Learning Tentative (JOLT) ¹	-	-	.26	.67
Knowledge Elaboration (KE)	6.64	6.45	3.91	3.56
Memorization (MEM)	.25	.72	.13	.40
Manipulate Representation (MR)	7.00	6.67	18.02	10.22
Monitoring Progress Towards Goal (MPG)	3.73	3.89	2.32	2.34
Monitoring Use of Strategies (MUS)	.55	1.07	.51	.98
Prior Knowledge Activation (PKA)	3.23	3.28	2.64	2.87
Plan (PLAN)	.20	.46	.43	.65
Recycle Goal in Working Memory (RGWM)	7.16	5.80	4.53	3.82
Read Notes (RN)	.82	1.37	.87	1.71
Representation Difficulty (RD)	1.82	1.85	.94	1.11
Re-reading (RR)	.25	.49	.83	1.54
Search (Search)	.61	1.26	.57	1.49
Select New Source of Information (SNIS)	23.48	7.88	21.43	7.38
Self-Knowledge Activation (SKA)	.82	1.83	1.21	1.73
Self-Questioning (SQ)	2.32	3.37	2.11	2.59
Sub-Goal (SG)	3.95	2.89	4.96	4.76
Summarizing (SUM)	10.25	6.71	9.60	5.47

Task Difficulty (TD)	.11	.32	.13	.40
Taking Notes (TN)	9.61	8.33	5.60	5.39
Time Monitoring (TM)	.66	1.01	.85	1.18
Time Planning (TP)	.57	.95	.89	1.34

---

¹Science Library-Only Code. ²History Library-Only Code

Table 7

Study 1 Multiplicative Change in Expected Micro-level SRL Count, with History as the Comparison Group.

<i>Micro-level SRL Process</i>	<i>Exp(b)</i>
CE+	2.00*
CE-	2.04*
EAC-	2.17*
ECRT-	2.36*
FOK+	1.54*
INT+	.43*
JOL+	4.21*
KE	.57*
RD	.51*
RGWM	.63*
RR	3.32*
TN	.59*

* $p < .05$ after Benjamini – Hochberg Correction (Benjamini & Hochberg, 1995).

Table 8

Study Two Descriptive Statistics

	Mean(<i>SD</i>)	Skewness(<i>SE</i>)	Kurtosis(<i>SE</i>)
<i>Science</i>			
Pretest	9.67(3.20)	1.18(.58)	2.06(1.12)
Posttest	15.67(4.19)	-.32(.58)	-1.44(1.12)
SRL Positive	62.40(24.85)	.48(.58)	-.61(1.12)
SRL Negative	3.60(2.97)	.88(.58)	.48(1.12)
<i>History</i>			
Pretest	8.00(7.76)	.42 (.58)	-1.74(1.12)
Posttest	23.93(4.35)	-.57(.58)	-.143(1.12)
SRL Positive	45.13(22.28)	.43(.58)	-.88(1.12)
SRL Negative	62.60(18.84)	.83(.58)	.96(1.12)

Table 9

Study 2 Dependent T-Tests between Pretests and Posttests

Digital Library	Mean difference	<i>df</i>	<i>t</i>	Cohen's <i>d</i>
Science	6.00	14	4.65*	.61
History	15.93	14	8.15*	.83

---

*p < .001

Table 10

Study 2 Data-Driven Aggregation Multiple Linear Regression Results, Standardized Regression Coefficients

	Science	History
<i>Predictor</i>		
Positive SRL	.41	.51
Negative SRL	-.10	-.13
Pretest	.08	.00
<i>F</i>	.847	1.31
<i>df</i>	3, 11	3, 11
Total R^2	.19	.26
Adjusted R^2	-.03	.06

* $p < .05$, ** $p < .01$

Appendix A

Classes, descriptions and examples of the macro- and micro-level SRL processes used to code students' regulatory behavior (based upon Azevedo et al., 2008; Greene & Azevedo, 2009). Micro-level SRL processes are grouped according to their superordinate macro-level SRL variable.

Micro-level processes	Description ^a	Student Example
<i>Macro-level variable: planning</i>		
Planning (PLAN)	Stating two or more sub-goals simultaneously or stating a sub-goal and combining it with a time requirement.	"First I'll look around to see the structure of environment and then I'll go to specific sections of the circulatory system."
Sub-goal (SG)	Learner articulates a specific sub-goal that is relevant to the experiment provided overall goal.	"I'm looking for something that's going to discuss how things move through the system."
Recycle goal in working memory (RGWM)	Restating the goal (e.g., question or parts of a question) in working memory.	"...describe the location and function of the major valves in the heart"
Time planning (TP)	Participant refers to the number of minutes remaining AND indicates whether a goal can be met during this time.	"How long is this article? Yeah, I have time to read it."
<i>Macro-level variable: strategy use</i>		
Comparing and contrasting (CC) ^b	Examining two separate representations or ideas (i.e., text, picture, simulation, etc.) to determine how they are similar and/or different.	"This picture shows blood leaving the heart through the aorta as red, but then the text says blood going through the pulmonary arteries is blue. There's something different between these two kinds of arteries."
Coordinating informational sources (COIS)	Using pointing or verbalizing the matching	"I'm going to put that [text] with the diagram."

---

	elements of two different representations, e.g., drawing and notes. Either representation can be in the environment or in participant's notes.	
Corroborating sources (COS) ^c	Comparing information from two separate sources, in the learning environment, to determine or verify their content as accurate.	"This says that the Blue Ridge Parkway went through North Carolina. I just want to check that with the map I saw."
Draw (DRAW)	Making a drawing or diagram to assist in learning.	"...I'm trying to imitate the diagram as best as possible."
Establishing chronology (EC) ^c	Participant determines when a historical event occurred; often in relation to another event but not necessarily.	"This picture must have been taken before the parkway was built, meaning before 1932." "This upper road was built after the vote on funding."
Historical perspective taking (HPT) ^c	Participant puts self in position of a historical figure; infers that figure's perspective, thinking, emotions; expresses understanding of that figure's decision making at that time.	"This letter seems harsh, I bet the writer was angry." "I can see why this person decided to vote against this proposal."
Hypothesizing (HYP)	Making a tentative conclusion by elaborating on what was just read, seen, or heard with prior knowledge. Making an informed guess (about content relevant to the task) based upon information either in the environment or from prior knowledge.	"The heat in this liquid is getting very high and the particles are moving faster, I bet it's going to over flow."
Inferences (INF)	Drawing a conclusion based	<i>"In this case, the transfusion</i>

---

	on two or more pieces of information that were read, seen, or heard in the hypermedia environment in same time period, roughly.	<i>will do the patient no good and also- and may even result in serious harm.</i> Probably because the antibodies are using energy to destroy it.”
Inferring source content (ISC)	Participant makes a guess as to the content available in a source.	“This link says it describes heat transfer, I bet that has some information on phase changes.”
Knowledge elaboration (KE)	Making a definitive conclusion by elaborating on what was just read, seen, or hear with prior knowledge.	...[Learner sees the diagram of the heart] and states “so the blood...through the...then goes from the atrium to the ventricle...and then...”
Manipulate Representation (MR)	Using pause, start, rewind, zoom, or other controls with a graphical representation.	“Hmmm...I’m going to zoom in on this letter [MR] to read it.”
Memorization (MEM)	Learner tries to memorize text, diagram, etc.	“One more time, aorta, the pulmonary artery, left atrium, aortic valve, mitral valve...”
Prior knowledge activation (PKA)	Searching memory for relevant prior knowledge either before beginning performance of a task or during task performance.	“the auricles were something on the test.”
Read notes (RN)	Learner reads over his/her own notes, drawings, etc.	“Let me read over these notes in the last few minutes.”
Re-reading (RR)	Re-reading or revisiting a section of the hypermedia environment.	“I’m reading this again.”
Search (SEARCH)	Searching the hypermedia environment.	Learner types in blood circulation in the search feature.

Self-knowledge activation (SKA)	The participant verbalizes that he or she is going to invoke a strategy because it is helpful to him/her personally. Or participant verbalizes that he/she is NOT going to invoke a strategy because it is NOT helpful to him/her. Or, participant says something about his/her own knowledge, beliefs, disposition, etc.	"I'm gonna take notes because that helps me remember things better."
Select new informational source (SNIS)	Using features of the hypermedia environment to access a new representation and/or a new section of the environment (clicking on hyperlinks, items in Table of Contents, back arrow). Only one SNIS will be coded, regardless of the number of clicks, until either the participant reads or enacts another SRL process.	[Learner double-clicks on the heart diagram to get a close-up of the structures.]
Summarization (SUM)	Verbally restating what was just read, inspected, or heard in the hypermedia environment.	"This says that white blood cells are involved in destroying foreign bodies."
Taking notes (TN)	Learner writes down information.	"I'm going to write that under heart."
<i>Macro-level variable: monitoring</i>		
Content evaluation (plus and minus) (CE+/-) ^d	Monitoring content relevant to goals. Learner states content is or is not useful toward reaching the goal.	"I'm reading through the info but it's not specific enough for what I'm looking for."
Expectation of adequacy of content (plus and minus) (EAC +/-) ^d	Expecting that a certain type of representation will prove either adequate or	"...the video will probably give me the info I need to answer this question" or "I

---

	inadequate given the current goal.	don't think this section on blood pressure will answer my question."
Evaluate content as relevant to task goal (plus and minus) (ECRT +/-) ^d	Statement that what was just read and/or seen is or is not relevant to an experimenter posed question or question from the learning tasks; i.e., recognition of relevance.	[Learner reads text]..."So, I think that's the answer to this question." "So, I do not think that's the answer to this question."
Emotion monitoring (EM)	Participant realizes that he/she is having an emotional response due to some aspect of the learning task.	"This text is frustrating me because it is so long."
Emotion Regulation (ER) ^e	Participant actively attempts to control emotional response to some aspect of the learning task.	"Okay, I need to calm down."
Feeling of knowledge (plus and minus) (FOK +/-) ^d	Learner is aware of having read something in the past and having some understanding of it, but not being able to recall it on demand or learner states this is information not seen before.	"..I recognize that from the pretest..." or "atherosclerosis – I never heard that word before."
Judgment of learning (plus and minus) (JOL +/-) ^d	Learner makes a statement that they understand what they've read or becomes aware that they don't know or understand everything they read.	"I get it" or "I don't know this stuff, it's difficult for me."
Judgment of learning tentative (JOLT) ^b	Participant has some understanding, is not sure that it is accurate, but indicates that s/he will proceed with that understanding until further evidence	"I'm not sure I understand this fully, but I'm going to go with it."

---

	confirms/disconfirms it.	
Monitor progress toward goals (MPG)	Assessing whether participant's previously-set goal and/or participant's own standard for understanding has been met.	"Those were our goals, we accomplished them."
Monitor use of strategies (MUS)	Participant comments on how useful a strategy is/was.	"Yeah, drawing it really helped me understand how blood flows throughout the heart."
Self-questioning (SQ)	The participant asks a question relevant to the task, but does not articulate a specific plan to investigate the answer. Indicates that the participant has recognized a gap in understanding.	"What is the difference between auricle and atrium?"
Time monitoring (TM)	Participant refers to the numbers of minutes remaining.	"There are a few seconds left."
<i>Macro-level variable: interest</i>		
Interest (plus and minus) (INT +/-) ^d	Learner has or does not have a certain level of interest in the task or in the content domain of the task.	"This stuff is interesting." OR "Boring."
<i>Macro-level variable: task difficulty and demands</i>		
Help seeking behavior (HSB)	Learner seeks assistance regarding either the adequacy of their understanding or their learning behavior, regardless of whether the instructions indicate that the experimenter/tutor will provide assistance.	"Why is the pulmonary artery colored blue?"

Representation Difficulty (RD)	Learner indicates the representation (i.e., picture, text, simulation) is not clear and/or unusable in general, regardless of one's learning goal.	"This small font makes the letter hard to read."
Task difficulty (TD)	Learner indicates one of the following: (1) the task is either easy or difficult, (2) the questions are either simple or difficult, (3) using the hypermedia environment is easier or more difficult than using a book.	"This is harder than reading a book."
<i>Miscellaneous</i>		
Tutor-Initiated Time Monitoring (TITM) ^e	Experimenter gives standardized time or prompt OR participant says "I'm done" and experimenter tells student number of minutes left.	"You have 8 minutes left."

---

^a All codes refer to what was recorded in the verbal protocols (i.e., read, seen, or heard in the environment and/or during discussions).

^b Micro-level process available for coding only within the science domain. All micro-level processes were available for coding in both domains unless denoted otherwise.

^c Micro-level process available for coding only within the history domain. All micro-level processes were available for coding in both domains unless denoted otherwise.

^d Plus and minus indicates that there are two separate codes. Plus is used when a participant notes the presence of the attribute and minus is used when the participant notes the absence of the attribute e.g., Content Evaluation (-) when the content is deemed not helpful by the participant.

^e Not included in the analysis for either study.

Appendix B Learning Tasks

The complete learning task prompt for the history domain was as follows: *“Imagine your teacher has given you an assignment: Using the resources available on the website Driving Through Time: The Digital Blue Ridge Parkway in North Carolina, learn all you can about the Blue Ridge Parkway and its origin. Let the following questions guide your exploration: What is the Blue Ridge Parkway and why was it built? Who was involved in its planning and construction? What cities, states, and regions were affected by the Parkway, and how? How did the Parkway affect various groups of people? What era of U.S. history saw the creation of the Parkway and how does the Parkway reflect that era? What different perspectives did different individuals or groups have about the Parkway, and what were some issues that caused conflict or controversy during the Parkway's construction?”*

The complete learning task prompt for the science domain was as follows: *“Imagine your teacher has given you an assignment: Your task is to use the site on your computer screen to provide a detailed explanation of what occurs during phase changes. An example of a phase change is when a substance like water in its solid state (ice) is heated and changes into a liquid state, and when heated more transitions finally to a gas state (steam). The following terms will be important in giving a well-formed description of the phase change process; temperature, heat, phase changes, molecular motion, kinetic energy, intermolecular forces and latent heat. Learn the material well enough so that you could pass a test that includes the following prompts: Explain the difference between temperature and heat. Describe how a heating curve graph shows changes in the temperature of a substance and how latent heat is represented on the graph. Explain how attractions between particles change during phase transformations and how the attraction changes relate to latent heat. List and describe the 3 major phases and the 4 major*

phase changes and how particle movement and attraction are related to it. Pressure is NOT an important concept for this learning task. How pressure changes affect temperature is not a learning goal for this activity.”

Appendix C
Science Knowledge Measure

1. Latent heat is ...
  - A. the energy produced by the affinity that the particles in the nucleus (i.e. protons and neutrons) of an atom have for each other.
  - B. the energy used to either hold or break the bonds that exist between particles in a substance
  - C. the attraction between particles in a substance that holds them together or near each other
  - D. the measure of how much energy is required to raise the temperature 1 degree for 1 gram of a substance
  - E. the energy produced by the affinity the electron particles have for the nucleus of a substance
2. Intermolecular Forces are...
  - A. the attractions that the particles in the nucleus (i.e. protons and neutrons) of an atom have for each other.
  - B. used to either hold or break the bonds that exist between particles in a substance
  - C. the attractions between particles in a substance that holds them together or near each other
  - D. the forces required to raise the temperature 1 degree for 1 gram of a substance
  - E. the attractions the electron particles have for the nucleus of a substance
3. Temperature is ....
  - A. the measure of how hot or cold a substance appears to be through touch
  - B. the measure of the amount of heat an object possesses
  - C. the measure of the attraction between particles in a substance that holds them together or close to each other
  - D. the measure of how fast the particles within a substance move
  - E. is the flow of energy internal energy from one substance to another
4. Kinetic energy is...
  - A. the opposite measure of potential energy
  - B. the measure of how fast a particle or object moves
  - C. the measure of the force needed to move an object
  - D. measured in meters per second
5. Heat is...
  - A. the forces required to raise the temperature 1 degree for 1 gram of a substance
  - B. the flow of internal energy from one substance to another
  - C. measured by Fahrenheit or Celsius
  - D. the measure of how fast the particles within a substance move
  - E. created usually by the motion of subatomic particles

Fill in the appropriate name of phase transitions described below.

6. ____ ____: liquid to solid

7. ____ ____: solid to liquid

8. ____ ____: liquid to gas

9. ____ ____: gas to liquid

10. Describe how the particles of solids, liquids, and gases move in relation to each other using the table below. Address particle attraction, movement, and spacing for each phase.

Solids	Liquids	Gases

Use the following figure to answer questions 11-16

11. During what portion(s) of the graph does the phase of the substance remain constant?
(Select one or more phases by circling the appropriate letters)

A   B   C   D   E

12. (Fill in the blank) At what temperature according to the graph is the phase of the substance changing from a solid to a liquid? _____

13. At what point(s) on the graph does evaporation and condensation occur?
(Select one or more phases by circling the appropriate letters)

A   B   C   D   E

14. (Fill in the blank) What phase(s)/state(s) of matter is the substance in at portion (D)? _____

15. At 0 degrees Celsius water can exist as partially a solid and a liquid, what is the major difference between the substances in its two phases at this point?

- A. There is no major difference
- B. The ice is colder than the water
- C. The particles in the ice have a stronger attraction to each other than those in the liquid water
- D. The particles in the liquid water are moving faster than those in the solid water ice

16. When measuring temperature which of the following best describes the phenomena being assessed?

- A. the force of the collisions between the particles in the substance
- B. the rate of travel at which the particles in the substance are moving
- C. the heat observed of the substance in comparison to a substance of known hotness
- D. the degree to which the substance is capable of heating another substance

Use the graph for question 17

17. During portion **(B)** on the graph heat in the form of energy is being added to the substance but the temperature is not changing, describe in as much detail as possible what is occurring? Make sure to explain where the heat energy is going. Also make sure to describe how latent heat is involved in this process.

Appendix D
History Knowledge Measure

1. What is the Blue Ridge Parkway?
  - a) A scenic recreational highway between two national parks
  - b) A North Carolina state park
  - c) A toll road running from Virginia to South Carolina
  - d) A commercial highway for farmers to transport their goods
2. Which of the following was a federal organization involved in its planning and construction?
  - a) The Federal Highway Commission
  - b) The National Park Service
  - c) The Little Switzerland Company
  - d) The Eastern Band of Cherokee
3. In what decade did construction on the Parkway begin?
4. In what decade was the Parkway completed?
5. What states does the Parkway run through?
  - a) North Carolina and Tennessee
  - b) Georgia and North Carolina
  - c) Tennessee and Virginia
  - d) Virginia and North Carolina
6. What state was considered for inclusion in the Parkway route but ultimately left out?
  - a) North Carolina
  - b) Tennessee
  - c) West Virginia
  - d) South Carolina
7. The Parkway is a product of which economic stimulus initiative?
  - a) The War on Poverty
  - b) The New Deal
  - c) The Marshall Plan
  - d) The Fourteen Points
8. Which one of the following groups benefited from the construction of the Parkway?
  - a) African-Americans
  - b) Poor farmers living along the Parkway
  - c) North Carolina tourism boosters and entrepreneurs
  - d) The U.S. timber industry

9. Which one of the following groups was negatively impacted by the Parkway?
- a) American tourists
  - b) Landowners along the Parkway route
  - c) The U.S. auto industry
  - d) Bicycle manufacturers
10. Which of the following caused a major controversy during the planning and construction of the Parkway?
- a) The routing of the Parkway over Grandfather Mountain
  - b) The National Park Service's treatment of African-Americans
  - c) The confiscation of Choctaw land for the Parkway's right-of-way
  - d) Partisan attacks on the constitutionality of the Parkway

11. Write what you know about the Blue Ridge Parkway and the historical context of its construction. Discuss the individuals/groups involved or affected, paying attention to different perspectives and conflicts between these people. Explain how the Parkway was significant to the era in which it was constructed.

***Do not forget to include information you wrote on the pre-test; the post-test will be graded on its own merits.^a

^a This note was only included on the post-test.

Domain-Specificity of Self-Regulated Learning Processing in Science and History

Jeffrey Alan Greene, Cheryl Mason Bolick, William P. Jackson, A. Michael Caprino,

Christopher Oswald, Megan McVea

University of North Carolina at Chapel Hill

CB#3500, Peabody Hall

Chapel Hill, NC 27599

*Email addresses: jagreene@email.unc.edu; cbolick@email.unc.edu; wpjackso@email.unc.edu;
caprino@email.unc.edu; coswald@live.unc.edu; mcvea@live.unc.edu*

Corresponding Author:

Jeffrey Alan Greene, Ph.D.

Associate Professor

113 Peabody Hall, CB#3500

School of Education

Learning Sciences and Psychological Studies Program

University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-3500

Contact Information

email: jagreene@email.unc.edu

office: 919-843-5550

fax: 919-843-2614

Authors' note: All authors' affiliation is with the School of Education, University of North Carolina at Chapel Hill; This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 103990. The authors were solely responsible for study design, data collection and analysis, the writing of the report, and the decision to submit the article for publication.