THE EFFECTS OF GOAL INSTRUCTIONS AND SOURCE REPRESENTATION SCAFFOLDS ON COLLEGE STUDENTS' INFORMATION EVALUATION AND ARGUMENTATION IN A GENERAL EDUCATION BIOLOGY COURSE

by

SO MI KIM

(Under the Direction of Michael J. Hannafin)

ABSTRACT

Online inquiry and argumentative writing constitute important academic tasks. Effective argumentation relies on evidence quality, grounded reasoning, and conceptual integration (Brem & Rips, 2000; Clark & Sampson, 2008), which can be enhanced by careful, deliberate information evaluation. This dissertation research aimed to examine mental tasks and conditions to activate seasoned information evaluation during argumentation based inquiry. Naïve task perceptions, lack of cue recognition, and cognitive load were addressed. Accordingly, source representation scaffolds and goal instructions were devised and implemented to test the effects in the college science classrooms. Source representation scaffolds aimed to model reflection over complex source properties while compensating cognitive capacity. The intervention included an annotation tool (treatment) and a checklist (control). Goal instructions were intended to induce critical task perceptions along with higher evaluation standards and efforts. The intervention included balanced reasoing goals (treatment) and persuasion goals (control).

Three manuscripts are included in this dissertation. Chapter 2 delineates the theoretical framework underpinning information evaluation scaffolds development and research. Chapter 3 reports findings from a longitudinal, quasi-experimental study that examines the effects of scaffolds on college students' self-reported information evaluation behavior change. Chapter 4 presents mixed methods research findings that examine whether and how source representation scaffolds and goal instructions influence information evaluation skills to improve argumentation quality. The results indicated that both goal instructions and source representation scaffolds treatment increased information evaluation behavior. However, goal instructions did not have direct effects on argumentation quality. Source representation demonstrated significant effects only when the annotation tool was combined with balanced reasoning goals and for students in heterogeneous knowledge groups. The findings together supported the synergistic integration of two scaffolding functionality, yet suggested addressing possible difficulties in using scaffolds and sustaining the effects in the complex classroom situations. The dissertation concludes with implications of the study and future research directions (Chapter 5).

INDEX WORDS: College Science, Information evaluation, Argumentation, Inquiry, Scaffolding, Source representation, Goal instructions, Task perceptions, Cognitive load, Group composition, Mixed Methods

THE EFFECTS OF GOAL INSTRUCTIONS AND SOURCE REPRESENTATION SCAFFOLDS ON COLLEGE STUDENTS' INFORMATION EVALUATION AND ARGUMENTATION IN A GENERAL EDUCATION BIOLOGY COURSE

by

SO MI KIM

B.A., Yonsei University, Republic of Korea, 1997

M.A., Yonsei University, Republic of Korea, 1999

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

THE EFFECTS OF GOAL INSTRUCTIONS AND SOURCE REPRESENTATION SCAFFOLDS ON COLLEGE STUDENTS' INFORMATION EVALUATION AND ARGUMENTATION IN A GENERAL EDUCATION BIOLOGY COURSE

by

SO MI KIM

Major Professor: Michael J. Hannafin

Committee: J. Michael Spector
Martha Carr
Marguerite Brickman

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
August 2014

DEDICATION

This dissertation is dedicated to God, whose promise has driven my journey, and to my family, whose trust has enabled the journey.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Michael Hannafin for his advice, patience, and support. I especially thank him for teaching me to appreciate dialogue between theory and practice across disciplines. His mentorship and academic philosophy have nurtured my funds of knowledge to guide research and practice now and forever.

I would also like to extend my appreciation to committee members, Drs. Michael Spector, Martha Carr, and Marguerite Brickman for their interest in my study and invaluable feedback that advanced my study. I give my special thanks to Dr. Marguerite Brickman for allowing me to conduct my study in her classroom. Partnership with her has been great encouragements for my journey.

Finally, I would like to thank colleagues and friends I have met at the University of Georgia. Due to my cohort group in the Learning, Design, and Technology program, I was not lonely throughout this journey. Also, my special gratitude goes to Katy and Lize, TAs in biology classrooms, who spent their invaluable year on supporting my research. Without their help, my study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In recent curricular reforms, undergraduate science educators have adopted online inquiry to bootstrap students' investigating and constructing scientific arguments (Krajcik, 2002; Lee et al., 2011; Linn, 2003; Narum, 2008; National Research Council, 1999). Scientific argumentation entails building valid and defensible answers to questions and problems when given information is incomplete or incompatible (Driver, Newton, & Osborne, 2000; Hirsch, 1989). Thus, scientific argumentation provides a basis for doing science in both academic and everyday contexts (Duschl & Osborne, 2002; Kuhn, 1993). A claim is built, evaluated, and gradually refined in light of newly gathered information through direct or often second-hand investigations (Brem, Russell, & Weems, 2001; Jiménez-Aleixandre & Erduran, 2007; Palincsar & Magnusson, 2001). The Web affords large enrollment science classes access to current, real-world issues and diverse perspectives and information sources (Krajcik, 2002; Narum, 2008; National Research Council, 1999). However, difficulties in building strong arguments with valid and reliable evidence have been identified (Kuhn, 1989; Larson, Britt, & Kurby, 2009; Takao & Kelly, 2003), which is especially true with largely unregulated sources of information on the Web (Apedoe & Reeves, 2006; Hannafin, Hannafin, & Gabbitas, 2009; Jones, 2001; Metzger, Flanagin, & Zwarun, 2003)

Information evaluation is a necessary but often unsupported and rarely activated focus when teaching argumentation (Brem et al., 2001; Iding, Crosby, Auernheimer, & Barbara Klemm, 2008; Wiley et al., 2009). Information evaluation involves activating key metacognitive skills needed to reflect over source features (e.g., the author, date, and abstracts) and to tag metadata (e.g. authority, currency, reliability, and convergence) involved in representing content across multiple sources (Perfetti, Rouet, & Britt, 1999; Stadtler & Bromme, 2007). For instance, information evaluation is a key factor determining information location (Tabatabai & Shore, 2005; Walraven et al., 2008). Effective information evaluation helps in assessing and discerning quality and coherence in collections of information sources (Mason, Boldrin, & Ariasi, 2009, 2010; Rouet, 2006), seeking elaboration and clarification (Tsai, 2004; Wu & Tsai, 2005), and constructing coherent and valid arguments (Bråten, Strømsø, & Britt, 2009; Wiley et al., 2009). However, these skills are rarely integrated into existing curriculum (Harris, 2007; Iding et al., 2008; Kuiper, Volman, & Terwel, 2005; Zhang & Quintana, 2012). Checklists, most commonly applied evaluation aid, may promote mechanical analyses (Harris, 2007; Meola, 2004; Metzger, 2007) ignoring complex mental tasks and conditions required for seasoned evaluation.

Three concerns have emerged: (1) task definition, (2) evaluation cues and criteria, and (3) metadata integration. First, college students tend to simplify their inquiry and focus on fact-finding (Alison & Michael, 2009; Asher, 2011). While college students report difficulties in narrowing focus, they rely on a small set of non-risky—for example, instructor proved—sources of information to solve the task in an efficient yet naïve way (Alison & Michael, 2009). Some students search for 'one best' answer even though all texts are open to dispute (Liang & Tsai, 2009; Tsai, 2004). Others orient their inquiry toward information to confirm rather than question their preconceptions (Kuhn, 1989; Nussbaum, Kardash, & Graham, 2005). Naïve beliefs and understandings related to a given task tend to reduce one's evaluation goals (Kienhues, Stadtler, & Bromme, 2011; Tsai, 2004; Wu & Tsai, 2005). For instance, students with low evaluation goals may fail to question whether important alternatives should be considered. Researchers have concluded that higher-level goal instructions encourage scrutiny and balanced reasoning over 'right' answers (Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Nussbaum et al., 2005).

Next, while college students often do not take information at face value, students base their evaluation on superficial cues such as titles, authors, or presence of statistics (Bråten et al., 2009; Britt & Aglinskas, 2002). Experts, on the contrary, take into consideration deeper levels of discipline specific features (e.g., evidence-claim structure) and seek corroboration from multiple sources when learning unfamiliar topics (Tsai, 2004; Wineburg, 1991). For instance, relational features such as agreeing, opposing, and stronger or weaker support may enable students to derive increasingly rich, coherent understanding from multiple information sources.

Lastly, effective information evaluation requires students to consider several mental representations simultaneously (e.g., navigation paths, site structures, content, and sources) (Brand-Gruwel, Wopereis, & Walraven, 2009; Brand-Gruwel, Wopereis, & Vermetten, 2005), which has been found to overtax working memory (Kalyuga, 2009; J.-F. Rouet, 2009; Stadtler & Bromme, 2008). Thus, students often do not voluntarily represent sources though they are able to clearly articulate their strategies when asked to do so (Eysenbach & Köhler, 2002; Stadtler & Bromme, 2004). Researchers have concluded that students can benefit from recording and organizing information features in concrete, explicit way that ease retention and update efforts of representation (Oestermeier & Hesse, 2000; Winne & Nesbit, 2009; Zhang & Quintana, 2012).

Research Purpose

Against this background, the current research program aimed to enhance student information evaluation and examine its impact on student argumentation during inquiry. The researcher studied required mental tasks, conditions, and approaches to enhance information evaluation behavior within an existing undergraduate biology curriculum by varying (1) source representation scaffolds and (2) goal instructions and examined the extent to which they influenced student evaluation and argumentation. Source representation scaffolds included a checklist (control) and a computer-based scaffold designed to externalize and integrate source representation within the argumentation process (treatment). Goal instructions included a persuasion goal ("persuade or convince the public of your claim") in the control condition and a balanced reasoning goal ("critically compare various claims" or "provide the reason why your counterarguments are wrong") in the treatment condition. It was assumed that, similar to other metacognitive processing, information evaluation support would guide students in defining information needs and collecting relevant evaluation cues while compensating for cognitive capacity to store multiple representations (Nelson & Narens, 1990; Perfetti et al., 1999; Winne & Hadwin, 1998).

Dissertation Overview

The dissertation consists of three ready to be published manuscripts. The first paper (Chapter 2) presents the theoretical framework underpinning information evaluation scaffolds development and research. The framework is developed following two interactive phases for the grounded design of instructional scaffolding (Hannafin, Hannafin, Land, & Oliver, 1997). First, a theoretical reference model is built to identify mental activities to be supported during argumentation. Then common student difficulties and empirical studies about associated scaffolding are reviewed to provide practical guidelines for classroom practice. Implications for future research are discussed at the end.

The second paper (Chapter 3) presents findings from a longitudinal, quasi-experimental study that examines the effects of differing instructional scaffolding—source representation scaffolds and goal instructions—on college students' information evaluation behavior (IEB) change over time. Using four self-report measures collected over an academic semester and their multilevel analyses, the study builds a model explaining the specific nature of change in IEB score. First, the study examines students' initial IEB status, growth rate, and their associations with individual differences—perceived task value (PTV) and working memory capacity to justify the need for goal instructions and source representation tools that compensate for cognitive capacity. Then the study examines the effects of varying scaffolds in terms of individual and combined effects as well as temporal change of the effects. Implications for practice, methodology, and future research are discussed at the end.

The third paper (Chapter 4) presents mixed methods research findings that examines whether and how source representation scaffolds and goal instructions influence the actual use of information evaluation skills to improve argumentation quality. The study first examines group and individual writing quality using Multivariate Analysis of Covariance (MANCOVA) to verify the individual and combined effects of scaffolds nested within varying group dynamics. Then students' task perceptions, source representation tool use experience, and cognitive load are qualitatively and quantitatively examined to understand the scaffolds use context. Students' task perception statements are thematically analyzed and clustered to summarize perception profiles. Source representation tool use experience is thematically analyzed with a focus on tool properties and tool use cases. NASA-Task Load Index score (Hart & Staveland, 1988) is computed to examine perceived cognitive load during the tool use. The study discusses final inferences based on the results from each strand at the end.

Finally, Chapter 5 presents a summary of key ideas from the three manuscripts. As the program of research has served for the initial understanding of information evaluation and its scaffolding approaches during inquiry-based argumentation, further research is required to devise, refine, and sustain acaffolding efforts in the dynamic classroom situations. Chapter 5 concludes with implications of the dissertation study, future research directions, and limitation.

Significance and Rationale

Thus far, the most common aid of information evaluation in the field has been a checklist that provides a list of evaluation criteria or questions. Though simple and efficient, a checklist does not make explicit evaluation cues and heuristics in a specific task context (e.g., argumentation) and does not address (meta) cognitive hurdles that face students, sometimes resulting in mechanical analyses (Harris, 2007; Meola, 2004; Metzger, 2007). Recent innovative studies focusing on supporting information evaluation have been reported across diverse areas (see for the review, Wiley et al., 2009; Wopereis & van Merrienboer, 2011). Despite promising findings, however, previous study durations have often been brief and conducted in laboratory settings. Criterion tasks are often contrived and do not reflect either the circumstances evident in everyday biology classes or immediate, pressing instructional concerns. Accordingly, research is needed to better support and develop students' scientific argumentation in real-world classroom contexts involving myriad information sources (c.f., Kim & Hannafin, 2011; Kuiper et al., 2005).

References

Apedoe, X. S., & Reeves, T. C. (2006). Inquiry-based learning and digital libraries in undergraduate Science Education. *Journal of Science Education and Technology*, *15*(5), 321–330.

Asher, A. D. (2011). Search magic: Discovering how undergraduates find information. Presented at the American Anthropological Association Annual Meeting, Montreal, Canada.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. doi:10.1016/j.chb.2004.10.005

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education*, *53*(4), 1207–1217. doi:10.1016/j.compedu.2009.06.004

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*(2/3), 191–213.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition & Instruction*, *20*(4), 485.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13–27. doi:10.1016/j.learninstruc.2007.12.003

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ (Clinical Research Ed.)*, *324*(7337), 573–577.

Hannafin, M. J., Hannafin, K., & Gabbitas, B. (2009). Re-examining cognition during student-centered, web-based learning. *Educational Technology Research and Development*, *57*(6), 767–785. doi:10.1007/s11423-009-9117-x

Hannafin, M. J., Hannafin, K. M., Land, S. M., & Oliver, K. (1997). Grounded practice and the design of constructivist learning environments. *Educational Technology Research and Development*, *45*(3), 101–117.

Harris, F. J. (2007). Challenges to teaching credibility assessment in Contemporary Schooling. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 155–180). Cambridge, MA: The MIT Press. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/dmal.9780262562324.155

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (Vol. Volume 52, pp. 139–183). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623869

Head, A., & Eisenberg, M. (2009). Lessons learned: How college students seek information in the digital age. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2009_finalv_YR1_12_2009v2.pdf

Hirsch, R. (1989). *Argumentation, information, and interaction: Studies in face-to-face interactive argumentation under differing turn-taking conditions*. Department of Linguistics, University of Göteborg.

Iding, M. K., Crosby, M. E., Auernheimer, B., & Barbara, E. (2008). Web site credibility: Why do people believe what they believe? *Instructional Science*, *37*, 43–63. doi:10.1007/s11251-008-9080-7

Jiménez-Aleixandre, M. P., & Erduran, S. (2007). Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education* (Vol. 35, pp. 3–27). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/content/v86q77727512075l/

Jones, B. D. (2001). Recommendations for implementing Internet inquiry projects. *Journal of Educational Technology Systems*, *30*(3), 271–291.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410.

Kienhues, D., Stadtler, M., & Bromme, R. (2011). Dealing with conflicting or consistent medical information on the web: When expert information breeds laypersons' doubts about experts. *Learning and Instruction*, *21*(2), 193–204. doi:doi: 10.1016/j.learninstruc.2010.02.004

Kim, M. C., & Hannafin, M. J. (2011). Scaffolding 6th graders' problem solving in technology-enhanced science classrooms: A qualitative case study. *Instructional Science: An International Journal of the Learning Sciences*, *39*(3), 255–282.

Krajcik, J. (2002). The value and challenges of using learning technologies to support students in learning science. *Research in Science Education*, *32*(4), 411–414. doi:10.1023/A:1022440215857

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689. doi:10.1037/0033-295X.96.4.674

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337.

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, *75*(3), 285–328.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, *77*(4), 339–365.

Lee, S. W.-Y., Tsai, C.-C., Wu, Y.-T., Tsai, M.-J., Liu, T.-C., Hwang, F.-K., … Chang, C.-Y. (2011). Internet-based science learning: A review of journal publications. *International Journal of Science Education*, *33*(14), 1893–1925.

Linn, M. C. (2003). Technology and science education: Starting points, research programs, and trends. *International Journal of Science Education*, *25*(6), 727–58.

Mason, L., Boldrin, A., & Ariasi, N. (2009). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607–633. doi:10.1007/s11251-008-9089-y

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, *5*(1), 67–90.

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy*, *4*(3), 331–344.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. doi:10.1002/asi.20672

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education*, *41*(3), 271–290. doi:10.1016/S0360-1315(03)00049-6

Narum, J. (2008). Transforming undergraduate programs in science, technology, engineering, and mathematics: Looking back and looking ahead. *Liberal Education*, *94*(2), 12–19.

National Research Council. (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology*. Washington, DC: National Academy Press.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–169). New York: Academic.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. doi:10.1037/0022-0663.97.2.157

Oestermeier, U., & Hesse, F. W. (2000). Verbal and visual causal arguments. *Cognition*, *75*(1), 65–104. doi:10.1016/S0010-0277(00)00060-3

Organisation for Economic Cooperation and Development. (2003). *The PISA 2003 assessment framework : Mathematics, reading, science and problem solving knowledge and skills*. Paris : OECD.

Palincsar, A. S., & Magnusson, S. J. (2001). The interplay of first-hand and secondhand investigations to model and support the development of scientific knowledge and reasoning. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 151–193). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum.

Rouet, J. (2006). *The skills of document use: From text comprehension to Web-based learning*. L. Erlbaum Associates.

Rouet, J.-F. (2009). Managing cognitive load during document-based learning. *Learning and Instruction*, *19*(5), 445–450. doi:16/j.learninstruc.2009.02.007

Stadtler, M., & Bromme, R. (2004). Laypersons searching for medical information on the web: The role of metacognition. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (p. 1638). Mahwah, NJ: Erlbaum.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, *2*(2-3), 191–210.

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the web search of laypersons. *Computers in Human Behavior*, *24*(3), 716–737. doi:10.1016/j.chb.2007.01.023

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research (07408188)*, *27*(2), 222–248. doi:10.1016/j.lisr.2005.01.005

Takao, A. Y., & Kelly, G. J. (2003). Assessment of evidence in university students' scientific writing. *Science & Education*, *12*(4), 341–363.

Tsai, C.-C. (2004). Information commitments in Web-based learning environments. *Innovations in Education and Teaching International*, *41*(1), 105–112.

Walraven, A., Brand-gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, *24*(3), 623–648. doi:doi: 10.1016/j.chb.2007.01.030

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. doi:10.1037/0022-0663.83.1.73

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York: Routledge.

Wopereis, I., & van Merrienboer, J. J. G. (2011). Evaluating text-based information on the World Wide Web. *Learning and Instruction*, *21*(2), 232–237.

Wu, Y.-T., & Tsai, C.-C. (2005). Information commitments: Evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, *21*(5), 374–385. doi:10.1111/j.1365-2729.2005.00144.x

Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, *58*(1), 181–196.

CHAPTER 2

SCIENCE ON THE WEB:

HOW DO COLLETE STUDENTS EVALUATE AND SUPPORT ARGUMENTS?[1]

[1] Kim, S. M., & Hannafin, M. J. Submitted to *Science Education*

Abstract

Though argumentation has been lauded as a rhetorical basis for doing science in both academic and everyday contexts, information evaluation has rarely been critically investigated in relation to the quality of argumentation. In this paper, we review theoretical underpinnings, related mental process, and results from empirical studies reporting challenges college students face during information evaluation. The analysis indicates that information evaluation is not dichotomous, but rather involves contextual, iterative meaning-making processes. College students typically simplify inquiry and evaluation tasks, report difficulties using advanced evaluation cues and criteria, and have limited cognitive capacities to retain, update, analyze, and connect diverse source representations. Associated scaffolding approaches are reviewed to provide practical guidelines for classroom practice.

Keywords: scientific argumentation, information evaluation, scaffolding

Introduction

Argumentation involves building valid answers to a question (Driver, Newton, & Osborne, 2000; Kuhn, 1993; Osborne, Erduran, & Simon, 2004). This process typically involves engaging a problematic situation in which students have incomplete or incompatible information (c.f., Gick, 1986; Hirsch, 1989). Students are challenged to resolve uncertainty in light of newly gathered information. The Web has served as a dynamic source of science information for the last decade (Brem, Russell, & Weems, 2001; Krajcik, 2002; Lee et al., 2011; National Research Council, 1999). However, efforts to build and support arguments using online science resources have resulted in myriad challenges for science educators. Students tend to simplify inquiries and encounter difficulties when weighing and integrating different resources (Head & Eisenberg, 2009; Hoffman, 1999; Hsieh & Tsai, 2011; Wallace, Kupperman, Krajcik, & Soloway, 2000; Walraven, Brand-gruwel, & Boshuizen, 2008). Kuhn (1993) notes that arguers need to "distance themselves from their own beliefs to a sufficient degree to be able to evaluate them as objects of cognition" (p. 331) using newly gathered information. Metacognitive skills become increasingly important as students interact with assorted, unregulated, and occasionally contradictory Web resources.

A critical metacognitive skill involves information evaluation, which enables reflection on source properties (e.g., author, date, source type, and thesis) across multiple resources (Brem et al., 2001; Perfetti, Rouet, & Britt, 1999; Wopereis & van Merrienboer, 2011). Information evaluation, thus, is key to locating critical sources (Tabatabai & Shore, 2005; Walraven et al., 2008), judging information coherence (Mason, Boldrin, & Ariasi, 2009, 2010; Rouet, 2006), seeking additional elaboration (Wu & Tsai, 2005), and constructing deep understanding (Bråten, Strømsø, & Britt, 2009; Wiley et al., 2009). Despite evidence of the benefits, educators have rarely integrated information evaluation skill into existing curricula (Harris, 2007; Kuiper, Volman, & Terwel, 2005; Zhang & Quintana, 2012). Information evaluation skill may be taken for granted (Iding, Crosby, Auernheimer, & Barbara, 2008) or treated as a mechanical rather than cognitive skill (c.f., Jonassen, 2010). Common instructional aids—a checklist, for example—reflect a librarians' bibliographical approach ignoring human factors (e.g., task perception and cognitive load) in a specific task context (Rieh, 2002). Indeed, educational researchers have only recently drawn attention to information evaluation (Lazonder & Rouet, 2008; Wopereis & van Merrienboer, 2011).

We present a framework for both understanding and supporting information evaluation during scientific argumentation. The framework is developed following two interactive phases for the grounded design of instructional scaffolding (Hannafin, Hannafin, Land, & Oliver, 1997). First, we examine the theoretical underpinnings of information evaluation to propose a reference model that identifies mental activities to be supported during argumentation. Then common student difficulties and empirical studies about associated scaffolding are reviewed to provide practical guidelines for classroom practice. Implications for future research are discussed at the end. This grounded-design approach rooted in theory enables the systematic analysis of differing challenges and guidelines necessary to promote critical reasoning processes. This review focuses specifically on college students—a population described as the first 'net generation' but perhaps assumed to be proficient and savvy Web users. Yet, research documenting student evaluation standards and application of Web-related academic tasks have proved inconclusive (Britt & Aglinskas, 2002; Mason et al., 2009). According to Kuhn (1989, 1993), students entering college tend to assimilate evidence to support preconceptions. Finally, we address the need to equip learners with the skills necessary to evaluate complex information to challenge rather than reify preconceptions.

Information Evaluation

Differing perspectives on information evaluation have been documented across fields from information science to text comprehension, and efforts to apply these views to educational problems have recently emerged (e.g., new literacy movement and information problem solving studies) (see for a review, Brand-Gruwel & Stadtler, 2011; Leu et al., 2008; Rieh & Danielson, 2007; Rouet, 2006). However, researchers' goals (e.g., information retrieval systems design, cognitive modeling, curriculum design, or instructional support), evaluation objects (a message, source, or medium), approaches (e.g., an bibliographical, human-computer interaction, information processing, or contextual approach), and media have varied (e.g., a single book, stand-alone system, or Web), complicating interpretation, synthesis, and implications of findings.

In this study, we define information evaluation as the knowledge and skill involved in *examining* and *using source properties* to derive meaning across resources. We, therefore, highlight how source properties are examined to select and process information (Bråten et al., 2009; Wineburg, 1991, 1998). We also classify information evaluation as largely metacognitive in nature, comprising (a) higher-order representation of source properties given a task and goal (Perfetti et al., 1999) and (b) control of subsequent cognitive efforts while defining, locating, and integrating information in need from multiple resources (Rieh & Hilligoss, 2008; Tabatabai & Shore, 2005). Since the seminal work of Flavell (1976, 1979), theorists have considered metacognition—active monitoring and control of cognition—as critical to intentional efforts to solve unfamiliar issues beyond one's knowledge threshold (Brown, 1987; Schraw, 1998; Winne & Hadwin, 1998). We concur with Hofer (2004), Kuhn (2000), and Mason, Boldrin, and Ariasi

(2010) that metacognition involves reflection on one's *knowledge sources* (e.g., "is this credible enough to support claim X?") as well as one's cognitive state, strategies, and task. This provides an analytic lens for complex psychological phenomena and enables synthesis across differing perspectives. In the following, we synthesize theories and empirical studies from diverse fields to examine how individuals represent sources and manage online information-seeking while referencing alternative representations.

Source Representation

Source representation is considered necessary to derive rich mental models across multiple resources (Bråten et al., 2009; Britt, Perfetti, Sandak, & Rouet, 1999). Researchers posit experts developed information-seeking strategies to represent unfamiliar problems (Gick, 1986; Simon, 1978). One key strategy involves information evaluation (Kirschenbaum, 1992; Tabatabai & Shore, 2005). Historians, for example, resolve historical interpretation problems using sourcing, contextualization, and corroboration heuristics (Wineburg, 1991, 1998). Sourcing involves examining source properties even before reading content; contextualization involves examining a source's temporal or spatial details; corroboration involves verifying source accuracy comparing more than two resources.

Text comprehension researchers have explored why source representation is essential to reading and learning of documents (Britt et al., 1999; Perfetti et al., 1999), and the implications also apply to learning with Web resources (Wiley et al., 2009). As depicted in Table 2.1, good readers construct a *documents model* combining *situation and intertext models.* The situation model is a semantic representation readers generate out of content (c.f., Dijk & Kintsch, 1983; Kintsch, 1991, 1998). The intertext model extends the situation model and consists of nodes representing individual document sources and their relational links. Because individuals vary in prior knowledge and reasoning skill, the same document might be represented differently—shallower or richer—in the situation model. Multiple, fragmented documents complicate reasoning processes further. Readers need to select and order information both within and across documents to create a global situation model. Moreover, each document has unique, incomplete, or conflicting base text, which often does not contain cross-references to guide comprehension across documents. Documents Model theory assumes that good readers develop reference (or intertext) models to recognize distinctions and relationships of separate, often conflicting documents. Several studies have supported this assumption across disciplines ranging from history (Bråten et al., 2009; Wineburg, 1998) to science (Stadtler & Bromme, 2007) and law (Stromoso & Braten, 2002).

Table 2.1

Components of the Documents Model

Level 1	Level 2	Level 3	Example
Documents Model	Global Situation Model	Local Situation Models from Individual Documents (Content)	Unique, Incomplete, and Conflicting Explanations
	Intertext Model (Source Representation)	Document Nodes (Source Properties)	Author, Date, Source Type, Purpose, and Thesis (e.g., claim summary)
		Source-to-Content Link	Written by, in, and about; thus Having Authority, Currency, and Topicality
		Source-to-Source Link	Agreeing, Opposing, or Caused by; thus reliable (strengthening), unreliable (weakening), or supplementing

How, then, do individuals represent sources? According to metacognitive theory, readers scan and judge cues (i.e., source properties) against certain criteria (e.g., reliability) (c.f., Winne & Hadwin, 1998). Documents Model theory posits that source representation processes are similar to those involved in semantic map construction. Readers fill-out (a) document nodes to mark a variety of source properties such as author, date, source type, purpose, and thesis. They then create links: (b) source-to-content links connecting a source and its content (e.g., "written by Dr. Browne") and (c) source-to-source links connecting separate sources in relation to content (e.g., "agreeing"). Though Document Model theory does not formulate evaluative judgment processes explicitly, literacy studies report *good readers* have acquired and use certain quality criteria to judge source properties (e.g., Pressley & Lundeberg, 2008). They presumably attach additional judgmental predicates to links enabling them to weigh content from differing sources (e.g., "written by Dr. Browne, thus having authority"; "agreeing, thus reliable").

The processes involved in linking, therefore, appear critical to establishing judgment criteria. Though quality has many underlying dimensions, information scientists note key distinctions between intrinsic and extrinsic judgment criteria (c.f., Knight & Burn, 2005; Saracevic, 2007; Schamber, Eisenberg, & Nilan, 1990). Early researchers, for example, employed bibliographical approaches (e.g., Smith, 1997; see for a review, Knight & Burn, 2005; Rieh, 2002) to link a source to content. Though such efforts have contributed to identifying separate criteria and designing feedback for information retrieval, some suggest that this approach examines only the intrinsic value, or *innate merit* of resources (e.g., currency, support, authority, and bias), which may have little relations to users' intentions (Rieh, 2002; Schamber et al., 1990). Others criticize the approach for encouraging mechanical, 'yes-or-no' evaluation (e.g., a checklist) (Harris, 2007; Meola, 2004). In later studies, quality was defined in terms of fit for needs in a specific context (Knight & Burn, 2005; Schamber et al., 1990). Extrinsic criteria have been suggested (Hjørland, 2012; Meola, 2004) to examine *relative strength and utility* for achieving a goal, or satisfying information needs. This approach seems related to source-to-source linking. Meola (2004), for example, suggested comparing sources to examine the depth and missing points or corroborating sources to verify information.

Critical source properties and criteria tend to vary across tasks and disciplines. A simple fact (e.g., "what is lactic acid?") can be located examining a topic fit on the search results; yet it is not the case for a complex task such as argumentation (e.g., "what causes burning sensation during exercise?") (c.f., Gerjets, Kammerer, & Werner, 2011). Setting and author information is needed to judge the credibility of historical narratives (Wineburg, 1998), but research methods and publisher information may be even more crucial in judging scientific arguments (Bråten et al., 2009). Familiarity with disciplinary discourse as well as domain knowledge exert domain-related influence over information evaluation (Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). Accordingly, research and evaluation training need to account for the influence of context (e.g., look-up search vs. argumentative writing) on evaluation in order to understand and support learner needs (Kuhlthau, 2008; Meola, 2004).

Interestingly, quality criteria for the same task can change over time. Wang and White (1999), for example, reported that relatively simple criteria (e.g., topicality) were used during the initial exploratory search. Additional criteria (e.g., depth) requiring close content examination were added during reading and citing stages. In Wineburg (1991) and Rieh and Hilligoss (2008), individuals initially examined intrinsic quality and then compared across sources to resolve contradictions between explanations. We speculate that evaluation focus change as one's cognitive state and information needs change (c.f., Saracevic, 2007; Schamber et al., 1990). For example, individuals dealing with unfamialir tasks initially tend to rely on intrinsic quality since the initial task demands rapid knowledge-base building rather than deep proecessing (c.f., Bowler, 2010). In contrast, external quality can be examined only using complex, subtle, contextual cues, which become apparent after reading and comparing resources (c.f., Knight & Burn, 2005).

Source representation, therefore, involves iterative meaning making relying on different criteria. Scientific arguments, for example, can be evaluated initially using publishing information; yet, in order to examine accuracy, source claims need to be compared. Accordingly, source and content representation require close interaction and subsequent adjustments (Britt et al., 1999; Perfetti et al., 1999). These processes become especially critical when examining dynamic and largely unfiltered information available online.

Control of Information-Seeking

Once individuals scan and judge source properties, the subsequent control process is hypothesized to reference updated representations to influence cognitive processes (c.f., Flavell, 1976; Winne & Hadwin, 1998). As illustrated in Figure 2.1, online information-seeking proceeds through multiple cycles in efforts to define, search, scan, process, and organize information in needs (Brand-Gruwel, Wopereis, & Walraven, 2009; Kuhlthau, 2004). Though many models delineate information evaluation as a sub-stage of sequential information-seeking (see for a review, Lazonder & Rouet, 2008), source representation is continuously updated and governs the entire process to complete a task (Rieh, 2002; Saracevic, 2007). Precisely how source representation influences information-seeking, however, is debatable. Some researchers have focused on a single effect of source representations, such as information location (e.g., Tabatabai & Shore, 2005) or comprehension (e.g., Perfetti et al., 1999), often involving a single document (e.g., Guthrie, 1988; McCrudden & Schraw, 2007) or pre-selected documents (Bråten et al., 2009; Wiley et al., 2009) . It is important, therefore, to examine information evaluation behaviors comprehensively rather than episodically. We analyze and synthesize diverse but complementary research across disciplines to examine how source representations evolve to influence search orientation, information location, integration, and finally reflection over the task outcome.

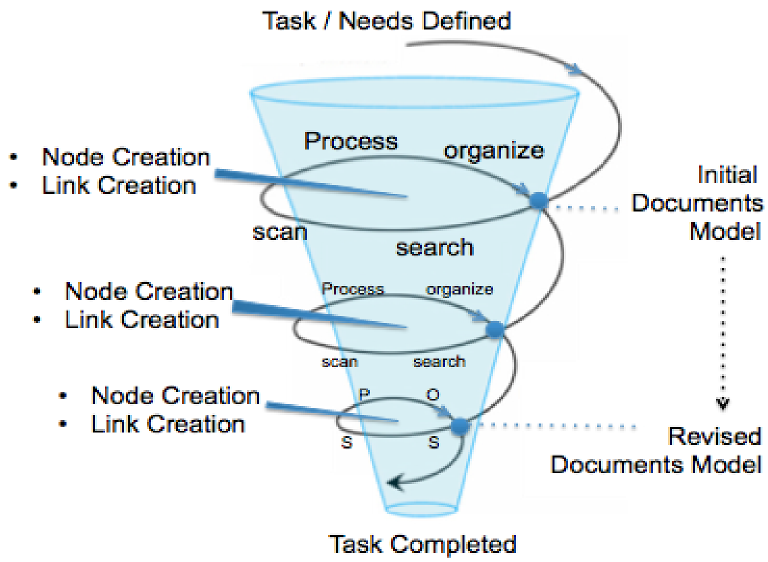

Figure 2.1. Mental process involved in information-seeking and evaluation.

Orienting information-seeking. Previous experiences with specific sources help to establish tacit evaluation standards and expectations for needed sources to guide information-seeking (Hogarth, 1980; Hughes, Wareham, & Joshi, 2010; Rieh & Hilligoss, 2008; Wu & Tsai, 2005). Rieh & Hilligoss (2008) found that college students engaged in predictive judgment about how and where to find the best information based on their task perceptions and prior experiences with specific sources. These predictive judgments influence subsequent information-seeking strategies. Tsai and his colleagues (Hsieh & Tsai, 2011; Liang & Tsai, 2009; Wu & Tsai, 2005) found that students with lower evaluation standards sought a single resource that aligned their initial perceptions. In contrast, college students with higher evaluation standards elaborated their previous search results and sought corroboration. Upon finding unverified information, they sought corroboration from other sources.

Locating information. When individuals address unfamiliar tasks, information needs are initially vague during searching and scanning (Taylor, 1967; Walraven et al., 2008). Uncertainty increases individuals' cognitive load as they rely on trial and error strategies to locate relevant resources (Kalyuga, 2011). Inefficient searchers, accordingly, tend to choose resources based on ease of access and surface cues (c.f., Kirschenbaum, 1992); efficient searchers, in contrast, allocate their cognitive resources selectively to valid cues to filter myriad resources (Tabatabai & Shore, 2005; Walraven et al., 2008). Tabtabai and Shore (2005), for example, examined coded verbatim data in order to explain how successful participants search for unknown facts. Findings revealed that the use of specific evaluation criteria was the most important factor. The judgment criteria used to locate information in the initial stages tended to be intrinsic (c.f., Rieh & Hilligoss, 2008; Wineburg, 1998) and focused on the innate merit of each resource. Establishing relevant intrinsic criteria, thus, serves as a minimum qualification to locate resources readily. The initially retrieved resources provide the foundation for extracting information and gradually reduce uncertainty (Taylor, 1968).

Integrating information. As the Documents Model theory suggests, source representations help to integrate information across resources (Bråten et al., 2009; Wiley et al., 2009). Wineburg (1998) compared how historians with different background knowledge explained Lincoln's views on race. The Lincoln expert consulted content knowledge while the other linked a series of documents referencing sources and their relationships. Source representations compensated for limited content knowledge. Recent studies have demonstrated explicit links between source representations and understanding. In Bråten et al. (2009), college students' awareness of trustworthy documents predicted successful comprehension tasks related to climate change. Wiley et al. (2009), who examined the influence of an evaluation tutorial on students' essays, reported that tutorial students performed superior on the test and demonstrated richer understanding with more conceptual nodes and links than those without the tutorial.

Reflecting over the task outcome. Though source and content representations involve incremental construction through successive cycles (Rouet, 2006), comparatively little attention has focused on the role of summative reflection on information evaluation. Metacognitive theory suggests that reflections may enhance short as well as long-term task performance (Pressley, 2002; Schraw, 1998). For example, even after reading, good readers evaluate the credibility of a reading text and think about how they are going to use it for a task (Pressley, 2002). Reflection presumably updates and strengthens source and content association (c.f., Bråten, Britt, Strømsø, & Rouet, 2011), supports reorganization of incompatible information (c.f., Russell, Stefik, Pirolli, & Card, 1993), and refines the final product. Reflection may also trigger the individual to reconsider the relevance of cues, criteria, or sources. The metacognitive knowledge again influences subsequent information-seeking process (Bowler, 2010; Rieh & Hilligoss, 2008; Schraw, 1998).

Information Evaluation During Argumentation

Argumentation builds on claim-evidence structure (Driver et al., 2000; Duschl & Osborne, 2002; Kuhn, 1993). A claim is a proposition or hypothesis that requires verification; evidence is data offered to verify a claim. Argumentation evolves as students gather and integrate fragmentary pieces of information, a process that necessitates continual information evaluation (Bell, 2000; Belland, Glazewski, & Richardson, 2008; Hirsch, 1989; van Eemeren, 1996; Voss & Means, 1991). Figure 2.2 depicts the flow between information and evaluation during argumentation in a hypothetical reference model. For simplification, we highlight core aspects of each stage, but this does not suggest that the process is sequential. In practice, several potential feedback effects may exist between and among argumentation, information, and information evaluation.

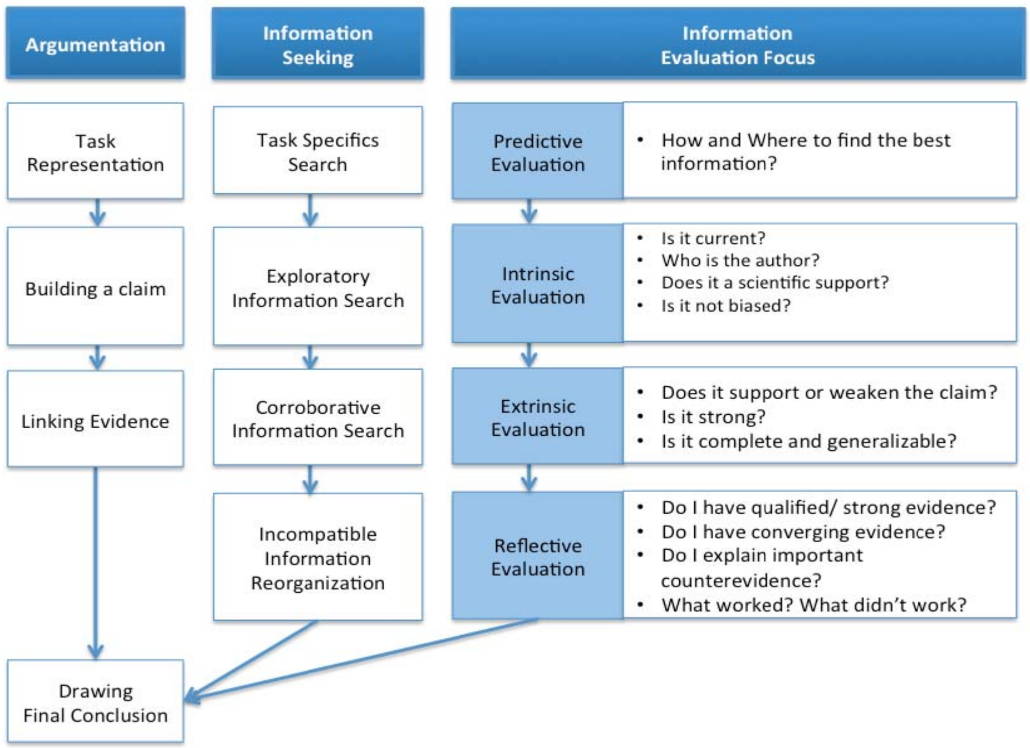

Figure 2.2. A hypothetical model of information evaluation during argumentation.

Task Representing and Predictive Evaluation

Given a task instruction, students mentally represent a task or question that induces argumentation and recall prior knowledge (Belland et al., 2008; Gick, 1986; Quintana, Zhang, & Krajcik, 2005). In doing so, they detect gaps between their current knowledge and the task requirements. Assume, for example, that students are to select a diet product and investigate differing claims about the safety. Searching for specifics of the task (e.g., "you need to examine ingredients, influence on weight loss, and side effects) helps to establish needs and guide their initial strategies. At this stage, students elicit evaluation standards to make a predictive evaluation about how and where to find relevant source information (Hogarth, 1980; Rieh & Hilligoss, 2008; Rieh, 2002). They may suspect that Wikipedia is unreliable but efficient and access it to formulate initial theories (c.f., Head & Eisenberg, 2010; Lim, 2009). Predictive evaluation orients the subsequent information-seeking behavior (c.f., Hughes, Wareham, & Joshi, 2010; Wu & Tsai, 2005). Some students may seek 'safe sources' and favor textbooks or instructor-selected resources (c.f., Alison & Michael, 2009; Bråten, Strømsø, & Salmerón, 2011). Others may have already evolved attitudes toward the inquiry topic ("diet products always involve side effects") and ignore or distrust resources with counter-evidence (c.f., Nussbaum & Kardash, 2005; Nussbaum, Sinatra, & Poliquin, 2008).

Exploratory Search for a Claim with Intrinsic Evaluation

Upon identifying knowledge gaps, individuals search for external sources to establish a tentative claim (Bell, 2000; Hirsch, 1989). The initial search is exploratory in nature and does not necessarily account for logical coherence or strength against a claim (Hirsch, 1989; Martindale, 1981). Rather, the goal is to capture important overlaps and narrow the focus of their effort (c.f., Russell, Stefik, Pirolli, & Card, 1993). Subsequently, students may identify document nodes (e.g., author, date, and purpose) and focus on source-to-source links to review intrinsic qualifications (e.g., authority and currency). For instance, students might locate several documents on the Web that report the effects of diet-pills containing Hydroxycut. They may attempt to filter qualified sources but do not necessarily compare the relative strengths of sources or examine counterevidence. While they tentatively build their claim of the influence of Hydroxycut on weight loss, their explanation is incomplete and fragmentary.

Linking Evidence with Extrinsic Evaluation

To support their claim, students may refine their search seeking evidence that increases the validity of their argument (Hirsch, 1989; Martindale, 1981; van Eemeren, 1996). Different sources are linked to the claim, compared, and integrated to refine their arguments. Students conduct extrinsic as well as intrinsic evaluations to judge the utility and veracity of different source evidence against their claim. For instance, students might question whether sources support or refute their claim, the significance of associated evidence, and whether sources provide complete and generalizable accounts (Mason et al., 2009; Meola, 2004). In effect, they weight different source evidence with their initial claim. For example, students may locate conflicting sets of evidence about the safety of Hydroxycut's key ingredient: caffeine. They compare explanations across sources and note consistencies as well as inconsistencies, which may support alternative explanations ("there is however another explanation according to Dr. so and so").

Reorganizing Incompatible Information with Reflective Evaluation

Through dialectical adjustment across differing claims and evidence, students reorganize incompatible information that does not fit into their claims (Russell et al., 1993; van Eemeren, 1996; Walton, 2008) to draw a final conclusion. Students reflect summatively on integrated evidence (Rieh & Hilligoss, 2008) to appraise products in light of their goals (c.f., Schraw, 1998). They judge the strength of their claim in the face of varied sources. They question the qualifications and strength of their evidence, the extent to which the evidence converges or diverges, and whether important counterevidence exists. For instance, students might have identified possible side effects of caffeine but recognize their evidence is too scarce or equivocal to dispute the initial claim. Evolved explanations therefore integrate multiple, balanced source information. Reflection helps to build a better documents model (c.f., Bråten, Britt, Strømsø, & Rouet, 2011; Wineburg, 1991). Concurrently, students can update their metacognitive knowledge about different cues they would need to examine (Bowler, 2010; Rieh & Hilligoss, 2008; Schraw, 1998).

Both argumentation and information evaluation evolves through dynamic interaction (Metzger, 2007; Rieh, 2002). As students interpret argumentation tasks, their definition set differing information needs and thus differing evaluation needs (Kuiper et al., 2005; Saracevic, 2007). To address those needs, students search, scan, process, and organize information with differing evaluation focus through multiple cycles (Tsai, 2004; Tabatabai & Shore, 2005; Perfetti et al., 1999). The results feed back to advance argumentation and refine information needs. Understanding and promoting effective information evaluation requires consideration of such differing needs and difficulties that arise at different points of task accomplishment (Rieh & Hilligoss, 2008).

Challenges During Information Evaluation

Unfortunately, most college students do not engage voluntarily in the full spectrum of information evaluation (e.g., Head & Eisenberg, 2010b). There are several barriers to engaging in meaningful information evaluation. In metacognitive terms, information evaluation involves eliciting standards in light of one's task and goals, collecting and comparing cues (i.e., sources properties) with standards, and interpreting the difference to match an appropriate action (i.e., re-reading, marking source distinctions, and seeking additional sources) (Pressley, 2002; Rouet, 2006; Schraw, 1998; Winne & Nesbit, 2009). However, researchers have continually noted that students (a) simplify their tasks, lowering evaluation standards and cognitive engagement, (b) report limited knowledge of evaluation criteria and cues, and (c) have limited cognitive capacity for retaining and connecting source representation to the situation models.

Simplifying Inquiries

College students often limit inquiries to basic fact-finding (Asher, 2011; Tsai, 2004) seeking efficiency (Head & Eisenberg, 2009). Naïve beliefs and understanding tend to lower evaluation goals and limit subsequent information elaboration (Kienhues, Stadtler, & Bromme, 2011; Wu & Tsai, 2005). In Tsai (2004)'s novice-expert comparison study, for example, college students considered online inquiry as a look-up search task rather than to explore or investigate (c.f., Marchionini, 2006). Asher (2011) reported that students' algorithmic search culture caused "twin problems of 'too little' and 'too much' information" (p.5) without elaboration. Similar results were reported in Alison and Michael's (2009) large-scale survey with 2,318 college students across the U.S., who made little use of opportunities to search freely on the Web. Students reported difficulty narrowing down a task, instead relying on a limited set of rick-free sources of information (e.g., instructor approved). Finally, Kuhn (1989) reported that students entering college assimilated evidence to support their claims without questioning whether alternative accounts existed. Students often replaced evidence with their hypothetical ideas (c.f., Brem & Rips, 2000).

Use of Superficial Cues and Evaluation Criteria

A further difficulty arises from reliance on superficial cues and evaluation criteria. Although college students are disinclined to accept information at face value, their evaluations are often limited to topical relevance and nominal authority (Alison & Michael, 2010) relying on titles, author affiliations, or presence of statistics (Britt & Aglinskas, 2002; Mason et al., 2009; Meola, 2004; Scholz-Crane, 1998). Evaluation, however, involves contextualized activity that requires attention to task contexts, deeper assessment of discipline-specific features (e.g., claim-evidence structure), and comparisons across different sources (Bråten, Britt, et al., 2011; Brem et al., 2001; Metzger, 2007). For instance, source properties such as stronger or weaker support are tagged to claims that enable students to derive increasingly rich, coherent understanding from multiple, incomplete resources.

Retaining, Updating, and Connecting Source Representation

Another important, but less explored, challenge relates to the cognitive load associated with deeper levels of analysis and evaluation. Effective information evaluation is not dichotomous, but rather involves contextual and iterative meaning making processes (Metzger, 2007; Rieh, 2002). Students need to retain and update document nodes and links with their evolving situation models. Information-seeking processes require simultaneous consideration of multiple alternative mental representations (e.g., navigation paths and site structures) (Brand-Gruwel, Wopereis, & Walraven, 2009; Brand-Gruwel, Wopereis, & Vermetten, 2005). These tasks may overwhelm working memory (Baddeley, 1986; Sweller, Ayres, & Kalyuga, 2011) while hampering and limiting needed meta-level source representation and regulation (Kalyuga, 2009; Rouet, 2009; Stadtler & Bromme, 2007). Stadtler and Bromme (2004) analyzed think-aloud protocols and reported that college students demonstrated only moderate metacognitive evidence while seeking health information. Similarly, in Eysenbach and Köhler (2002), adults who searched for health information rarely activated evaluation strategies but they were able to articulate their strategies when asked to do so. College students, though they scan information quality for course work perceived as important (Head & Eisenberg, 2010b), appear to experience difficulties when combining multiple sources to reorganize incompatible information.

Scaffolding Information Evaluation During Argumentation

Information evaluation, therefore, involves contextual, iterative, meaning-making processes (Metzger, 2007; Rieh, 2002) built on interactions among task environments, student, and information (Lazonder & Rouet, 2008). Recent reviews of college students’ access and use of Web resources suggests that the ‘net generation’ lacks proficiency in evaluating issues related to academic studies (Rieh & Hilligoss, 2008). Researchers and designers have applied scaffolding to reduce discrepancies by providing temporary support in real-task contexts (Pea, 2004; Stone, 1998; Wood, Bruner, & Ross, 1976). We review scaffolding approaches associated with specific difficulties during information evaluation.

Orienting Goals and Inquiries

Goals activate mental schema that direct attention to relevant information and direct effort (Song, 2005; Locke & Latham, 2002). Goals serve as standards that guide monitoring and regulation of learning processes (Jiang & Elen, 2011; Winne & Hadwin, 1998). In a similar vein, goals guide evaluation efforts (Stadtler, Scharrer, & Bromme, 2011). It is critical to address naïve goals in order to determine local or global coherence among sources. A number of researchers have scaffolded student goal-setting using different tasks, adjunct questions, goal instructions (i.e., directions about what to accomplish), or prompts that help with goal articulation (Cerdán & Vidal-Abarca, 2008; Rouet, Vidal-Abarca, Erboul, & Millogo, 2001).

In order to promote claim-evidence structure and evaluative evidence, writing argumentation has been suggested to facilitate concept integration and causal connections more than writing narratives alone (Wiley & Voss, 1999; Wiley et al., 2009). However, the effects of argumentation tasks have occasionally proven inconclusive, especially when students address conflicting or incomplete information sources. In Stadtler et al. (2011), for instance, college students were provided pre-selected websites that contained mutually conflicting information and asked to write an essay to inform peers about cholesterol levels. No significant differences were reported across task conditions regarding the number of source references to conflicting information. In Wiley et al. (2009), the argumentation task alone did not trigger across-site comparisons. These results are similar to Kuhn's (1989, 1993) observations that college students assimilate evidence to support their claims while overlooking conflicting arguments. One caveat, however, is that intended goals may not necessarily be aligned with perceived goals (Jiang & Elen, 2011; Locke & Latham, 2002; Winne, 1983). Therefore, scaffolding needs to support students' perceived goals.

Social sciences researchers have examined goal instructions that address students' perception and beliefs about the nature of argumentative tasks (Ferretti, MacArthur, & Dowdy, 2000; Nussbaum & Kardash, 2005; Page-Voth & Graham, 1999). Nussbaum and Kardash (2005), for instance, made necessary rhetorical moves (e.g., rebuttal) explicit in the assigned goal statement to address students' my-side bias (i.e., the tendency to favor their own position during argumentation). Students were asked to write an essay about the harmful effects of TV violence given three different goals: base line ("explain your claim"), reasoning ("provide as many reasons as possible"), and rebuttal ("provide the reason why your counterarguments are wrong"). Results indicated that students wrote more counter-arguments and rebuttals given rebuttal-goal instructions than both the reasoning or support your claim conditions. The authors cautioned against general persuasion goals (e.g., "convince or persuade your friend about your claim") without also addressing students' perception of the task. Though the task in Nussbaum and Kardash's study did not involve the free information search, their caution seems especially relevant to online inquiries involving multiple resources. Unless students intentionally seek alternative sources to verify existing sources, they are apt to mistakenly process incomplete information.

Increasing Cue Recognition

Evaluation involves comparing observations with goals or standards (Azevedo, Guthrie, & Seibert, 2004; Winne, 2001). Both internal and external cues can trigger evaluation. Information evaluation, for instance, incorporates representation of external cues (e.g., source features) (Perfetti et al., 1999). Such cues provide feedback that is critical to effective monitoring. Unaware of complexity (Dunlosky, Rawson, & Hacker, 2002; Griffin, Wiley, & Thiede, 2008; Manlove, Lazonder, & Ton de Jong, 2007), however, students may rely on superficial cues. To increase the awareness and use of deep cues, direct instruction, expert modeling, and prompting have been employed to externalize otherwise implicit cues.

When the cues are applicable across tasks and independent of content, direct instruction provides a simple and efficient approach to training cues (c.f., Osman & Hannafin, 1992). For example, checklists, commonly used during information evaluation, extract and formalize evaluation criteria. The criteria, however, typically focus on internal qualifications (e.g., currency, authority, and purpose) regardless of the task or disciplinary context of documents. Though simple and efficient, researchers have criticized checklist approaches as focusing on highly context-dependent, meaning-making processes (Harris, 2007; Meola, 2004; Metzger, 2007).

On the contrary, relevant criteria also reflect discipline, task, or situation-specific contexts (Bråten et al., 2009; Gerjets et al., 2011; Saracevic, 2007). Information evaluation requires students to examine both overt (e.g., source features) and covert (e.g., content or how content has been created) aspects of documents that, when combined and inter-related, contribute to overall impressions. Credibility cannot be judged solely based on the target source's surface features (e.g., the author); rather, assessing credibility involves comparisons across sources with respect to completeness of explanations or rigor in methodologies. Students can internalize evaluation strategies and detect authentic, contextual, and semantic cues as well as surface features (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011).

Accordingly, facility with contextually-based approaches to information evaluation require attention (c.f., Lin, 2001; Manlove et al., 2007). Cognitive apprenticeships (c.f., Collins, Brown, & Newman, 1989), for example, make explicit an expert's reasoning during authentic tasks; external representations provide specific comparison points or cues. In Wiley et al.(2009)'s study, college students were provided expert models (i.e., site rakings and rationales) while engaged in a site evaluation task. Prompting may help students to internalize important questions and direct attention to associated cues. Britt and Aglinskas (2002) developed computer-supported learning environments in which students were presented historical problems and associated documents. Students were prompted to fill out a structured note card while conducting a reading-to-write task.

Successful approaches rely on what Brown et al. (1983) described as "informed training" (p.163)" plus "self-control" (p. 166) in which students are explicitly informed of specific conditions or cues that signal successful performance as they engage opportunities to manage their performance using those cues. Students are, in turn, encouraged to articulate why and how and which strategies and cues improve their performance. Comments from peers, instructors, experts, or computer-automated tracking records feed back to students to correct their interpretation of cues (Pressley et al., 1992; Winne & Nesbit, 2009). This fosters a self-sustaining cycle of metacognitive learning (Borkowski, Carr, & Pressley, 1987; Pressley et al., 1992).

Online Annotation Tools for Optimizing Cognitive Load

Even when aware of cues associated with information quality, students often do not voluntarily activate evaluation strategies due to perceived cognitive load and required effort (Eysenbach & Köhler, 2002; Rouet, 2009; Stadtler & Bromme, 2004). Annotation tools, or note-taking technology, may help students to generate meaning from online resources while minimizing cognitive load (c.f., Lee, Huang, Liu, Wang, & Hsu, 2011). Annotations are notes appended to traditional documents, including highlights, bookmarks (location information), tags (keywords), source information, and external representations attached to online documents (Bottoni, Levialdi, & Rizzo, 2003; Marshall, 1997; Suwa & Tversky, 2002). External representations allow external access to information, thus compensating for limited working memory throughout during information-seeking (Kobayashi, 2006). While annotation provides externalized memory aids, students can extract, reorganize, and collaborate on important source features and content (Lee, 2004b).

To optimize cognitive load efficiency when annotating, it is important to consider four principles: integration, contextualization, decomposition, and automation. Information-seeking tasks concurrently involves linking tools (e.g., search, evaluation, and organization tools) with representations (e.g., key words, paths, sources, and content) (Belland et al., 2008; Lee, 2004a; Miyata & Norman, 1986; Zhang & Quintana, 2012). The use of excessive tools and representations can inadvertently cause the split attention effect (Chandler & Sweller, 1992), where focus is divided rather than converged. Integrated workspaces, therefore, are needed to support linking with artifacts to reduce the extraneous effort involved in shifting between activities (Zhang & Quintana, 2012).

External representations also serve as information retrieval cues (Lee, 2004a; van der Pol, Admiraal, & Simons, 2006; Wolfe, 2008). When annotation tools are embedded within cognitive task context (e.g., argumentation), retrieval effect is maximized while reducing redundancy involved in revisiting documents or rebuilding task context. Van der Pol et al. (2006) embedded an annotation interface into a group discussion board to enable students to share comments about assigned reading. The interface helped students easily retrieve the context for a particular discussion threads. Lee (2004a) embedded annotations into a concept mapping process. The semantic cues afforded by concept maps increased the reusability of annotations while decreasing the cognitive effort needed to retrieve associated information.

Decomposition of activities becomes critical when a task spans extended, complex learning activities over time (c.f., Wood, Bruner, & Ross, 1976). Source representations during inquiry tasks involve iterative meaning-making that requires continual updates involving aspects of different documents. Britt and Aglinskas (2002) decomposed evaluation criteria using different note-taking tabs. The separation not only made expert strategies visible but also reduced the cognitive effort required to combine the processing of multiple representations.

Finally, automated supports are designed to help students to concentrate on important cognitive activities (Kim & Reeves, 2007; Quintana, Zhang, & Krajcik, 2005; Zhang & Quintana, 2012). Zhang and Quintana (2012), for instance, automatically saved log files to record students' search history, including search words, URLs, and results. This logging function freed students from having to save and store search results and helped them focus on synthesis and reflection.

Implications for Future Studies

Both formal science education and everyday science literacy rely increasingly on the Internet for information. However, student proficiency in information evaluation is rarely addressed. Our framework provides a much-needed basis for further analyzing student activities and identifying research needed to improve the quality of student argumentation.

First, to improve information evaluation to support argumentation, it is important to clarify the associated mental process and to determine how resources enhance, extend or limit student reasoning. The proposed framework building upon both direct and indirect results, provides a theoretically and inductively derived model to refine the processes associated with information evaluation. Empirical studies are needed to clarify the argumentation processes involved and to specify student difficulties during information evaluation.

Second, information evaluation requires students to define a task, collect appropriate evidence, and interpret cues in light of task directives and goals. Again, interpretation of cues references individual outcome expectancies and external feedback, promoting awareness of rigorous conditional knowledge as to when, where, and why certain strategies and cues are employed (Borkowski et al., 1987; Pressley et al., 1992). Scaffolding is needed to span the entire process in order to promote self-sustaining metacognitive learning (c.f., Zimmerman & Tsikalas, 2005). For example, a student's task definition defines and influences subsequent search for and evaluation of information. Without addressing student perceptions of a given task, scaffolding might support only mechanical application of imposed activities, as is often the case for other metacognitive or higher-order cognitive activities (Sharma & Hannafin, 2007; Zimmerman & Tsikalas, 2005).

Third, information evaluation includes dynamic interactions among students, information, and task variables (Kuiper et al., 2005; Lazonder & Rouet, 2008; Rieh & Hilligoss, 2008). Many current approaches ignore such dynamics in the context of real-world tasks (Wolfe, 2008). Research is needed to examine effects across different variables. For instance, student characteristics (e.g., levels of prior domain and discourse knowledge), task requirements (e.g., fact finding, summary, and argumentation), and the nature of information (e.g., pre-selected information or free-searched information) appear likely to yield different effects. Future studies should provide stronger rationale as well as empirical support for various scaffolding strategies.

Fourth, dynamic interactions among students, information, and tasks also provide reference models needed to track learning progression and subsequently adapt scaffolding designs (Berland & McNeill, 2010; Kuiper et al., 2005). Depending on students' developmental progression, a task might be initiated via a simple inquiry with a few possible answers but evolve into a complex inquiry with multiple alternatives, necessitating further support and/or rebuttal (Berland & McNeill, 2010). An instructor might introduce tasks using pre-selected resources that include task-relevant as well as irrelevant information for evaluation training purposes. After initial rounds, students can be encouraged or required to pursue individual searches to confirm, refute or qualify working theories.

Five, many information and library sciences studies have focused on systems designed to simplify the search process. Education perspectives, however, focus on transforming learning experiences—promoting meaning-making process over simple identification (Moraveji, Morris, Morris, Czerwinski, & Riche, 2011; Quintana, Shin, Norris, & Soloway, 2006). For instance, annotation tools designed properly to promote evaluation considering argumentation tasks might support individual interpretation and encoding of evaluation cues while minimizing extraneous cognitive load (Lee, 2004; Wolfe, 2008) and enhancing argumentation quality (Wiley et al., 2009)

Finally, evaluation is not dichotomous, nor does it occur at a single point in time. Rather, information evaluation involves iterative, meaning-making processes. Information evaluation starts with predictive judgment and evolves as multiple cues are collected over time. Better understanding the nature of this process might help to explain the largely inconclusive research findings regarding college students' evaluation skills (Britt & Aglinskas, 2002; Mason et al., 2009). As Wiley et al. (2009) suggested, directive evidence confirming the effects of information evaluation on student performance is needed. Accordingly, we need to track and monitor changes in real-time student evaluation patterns over extended tasks.

REFERENCES

Asher, A. D. (2011). Search magic: Discovering how undergraduates find information. Presented at the American Anthropological Association Annual Meeting, Montreal, Canada.

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Bell, P. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, *22*(8), 797–817. doi:10.1080/095006900412284

Belland, B. R., Glazewski, K., & Richardson, J. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research & Development*, *56*(4), 401–422. doi:10.1007/s11423-007-9074-1

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 775–793. doi:10.1002/sce.20402

Borkowski, J. G., Carr, M., & Pressley, M. (1987). "Spontaneous" strategy use: Perspectives from metacognitive theory. *Intelligence*, *11*(1), 61–75. doi:16/0160-2896(87)90027-4

Bottoni, P., Levialdi, S., & Rizzo, P. (2003). An analysis and case study of digital annotation. In N. Bianchi-Berthouze (Ed.), *Databases in Networked Information Systems* (Vol. 2822, pp. 216–230). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://www.springerlink.com/content/u1cu2mx7cexjf01h/

Bowler, L. (2010). A taxonomy of adolescent metacognitive knowledge during the information search process. *Library & Information Science Research*, *32*(1), 27–42. doi:16/j.lisr.2009.09.005

Brand-Gruwel, S., & Stadtler, M. (2011). Solving information-based problems: Evaluating sources and information. *Learning and Instruction*, *21*(2), 175–179. doi:doi: 10.1016/j.learninstruc.2010.02.008

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. doi:10.1016/j.chb.2004.10.005

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education*, *53*(4), 1207–1217. doi:10.1016/j.compedu.2009.06.004

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48–70. doi:10.1080/00461520.2011.538647

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning & Instruction*, *21*(2), 180–192. doi:10.1016/j.learninstruc.2010.02.002

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science*, *24*, 573–604. doi:10.1016/S0364-0213(00)00033-1

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*(2/3), 191–213.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition & Instruction*, *20*(4), 485.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso.* (pp. 209–233). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weiner & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Makham (Eds.), *Handbook of child psychology: Vol. III. Cognitive development* (pp. 77 – 166). New York: Wiley.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209–222. doi:10.1037/0022-0663.100.1.209

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233–246. doi:10.1111/j.2044-8279.1992.tb01017.x

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press,.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Dunlosky, J., Rawson, K. A., & Hacker, D. J. (2002). Metacomprehension of science text: Investigating the levels-of-disruption hypothesis. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension.* (pp. 255–279). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ (Clinical Research Ed.)*, *324*(7337), 573–577.

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, *92*(4), 694–702. doi:10.1037/0022-0663.92.4.694

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–236). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. doi:10.1037/0003-066X.34.10.906

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. doi:doi: 10.1016/j.learninstruc.2010.02.005

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, *21*(1 & 2), 99–120.

Griffin, T., Wiley, J., & Thiede, K. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*(1), 93–103.

Guthrie, J. T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly*, *23*(2), 178–199. doi:10.2307/747801

Hae-Deok Song. (2005). Motivating ill-structured problem solving in a Web-based peer-group learning environment: A learning-goal perspective. *Journal of Educational Computing Research*, *33*(4), 351–367.

Hannafin, M. J., Hannafin, K. M., Land, S. M., & Oliver, K. (1997). Grounded practice and the design of constructivist learning environments. *Educational Technology Research and Development*, *45*(3), 101–117.

Harris, F. J. (2007). Challenges to teaching credibility assessment in contemporary schooling. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 155–180). Cambridge, MA: The MIT Press. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/dmal.9780262562324.155

Head, A., & Eisenberg, M. (2009). Finding context: What today's college students say about conducting research in the digital age. University of Washington's Information School. Retrieved from http://projectinfolit.org/pdfs/PIL_ProgressReport_2_2009.pdf

Head, A., & Eisenberg, M. (2010a). How today's college students use Wikipedia for course-related research. *First Monday*, *15*(3). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2830

Head, A., & Eisenberg, M. (2010b). Truth be told: How college students evaluate and use information in the digital age. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2010_Survey_FullReport1.pdf

Hirsch, R. (1989). *Argumentation, information, and interaction: Studies in face-to-face interactive argumentation under differing turn-taking conditions*. Department of Linguistics, University of Göteborg.

Hjørland, B. (2012). Methods for evaluating information sources: An annotated catalogue. *Journal of Information Science*, *38*(3), 258–268. doi:10.1177/0165551512439178

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist*, *39*(1), 43–55.

Hoffman, J. L. (1999). Information-seeking strategies and science content understandings of sixth grade students using on-line learning environments. University of Michigan.

Hogarth, R. M. (1980). *Judgment and choice : The psychology of decision*. John Wiley & Sons Inc.

Hsieh, Y.-H., & Tsai, C.-C. (2011). The relationships among science-related major students' information commitment, mental load and mental effort. In T. Hirashima (Ed.), *Proceedings of the 19th International Conference on Computers in Education*. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education.

Hughes, B., Wareham, J., & Joshi, I. (2010). Doctors' online information needs, cognitive search strategies, and judgments of information quality and cognitive authority: How predictive judgments introduce bias into cognitive search models. *Journal of the American Society for Information Science & Technology*, *61*(3), 433–452. doi:10.1002/asi.21245

Iding, M. K., Crosby, M. E., Auernheimer, B., & Barbara, E. (2008). Web site credibility: Why do people believe what they believe? *Instructional Science*, *37*, 43–63. doi:10.1007/s11251-008-9080-7

Jiang, L., & Elen, J. (2011). Why do learning goals (not) work: a reexamination of the hypothesized effectiveness of learning goals based on students' behaviour and cognitive processes. *Educational Technology Research and Development*, *59*(4), 553–573. doi:10.1007/s11423-011-9200-y

Jonassen, D. H. (2010). *Learning to solve problems: A Handbook for designing problem-solving learning environments*. New York, NY: Routledge.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410.

Kalyuga, S. (2011). Informing: A cognitive load perspective. *Informing Science*, *14*, 33–45.

Kienhues, D., Stadtler, M., & Bromme, R. (2011). Dealing with conflicting or consistent medical information on the web: When expert information breeds laypersons' doubts about experts. *Learning and Instruction*, *21*(2), 193–204. doi:doi: 10.1016/j.learninstruc.2010.02.004

Kim, B., & Reeves, T. (2007). Reframing research on learning with technology: In search of the meaning of cognitive tools.

Kintsch, W. (1991). The role of knowledge in discourse comprehension: A Construction-Integration Model. In *Text and Text Processing* (pp. 107–153). Amsterdam: North-Holland.

Kintsch, W. (1998). *Comprehension : A paradigm for cognition*. New York: Cambridge University Press.

Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology*, *77*(3), 343–352. doi:10.1037/0021-9010.77.3.343

Knight, S., & Burn, J. (2005). Developing a framework for assessing information quality on the World Wide Web. *Informing Science*, *8*, 159–172.

Kobayashi, K. (2006). Combined effects of note-taking/-reviewing on learning and the enhancement through interventions: A meta-analytic review. *Educational Psychology*, *26*(3), 459–477.

Krajcik, J. (2002). The value and challenges of using learning technologies to support students in learning science. *Research in Science Education*, *32*(4), 411–414. doi:10.1023/A:1022440215857

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services* (2nd ed.). Westport, CT: Libraries Unlimited.

Kuhlthau, C. C. (2008). From information to meaning: Confronting challenges of the twenty-first century. *Libri: International Journal of Libraries & Information Services*, *58*(2), 66–73.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689. doi:10.1037/0033-295X.96.4.674

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind.* (pp. 301–326). Hove England: Psychology Press.

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, *75*(3), 285–328.

Lazonder, A. W., & Rouet, J.-F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, *24*(3), 753–765. doi:10.1016/j.chb.2007.01.025

Lee, S. W.-Y., Tsai, C.-C., Wu, Y.-T., Tsai, M.-J., Liu, T.-C., Hwang, F.-K., … Chang, C.-Y. (2011). Internet-based science learning: A review of journal publications. *International Journal of Science Education*, *33*(14), 1893–1925.

Lee, W.-C., Huang, S.-P., Liu, Y.-C., Wang, S.-R., & Hsu, W.-C. (2011). An Online Summary Writing System Combining with Concept Mapping and Annotation Sharing. In M. Chang, W.-Y. Hwang, M.-P. Chen, & W. Müller (Eds.), *Edutainment Technologies. Educational Games and Virtual Reality/Augmented Reality Applications* (Vol. 6872, pp. 392–396). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://www.springerlink.com/content/t80864v375881h76/

Lee, Y.-J. (2004a). Concept mapping your Web searches: a design rationale and Web□enabled application. *Journal of Computer Assisted Learning*, *20*(2), 103–113. doi:10.1111/j.1365-2729.2004.00070.x

Lee, Y.-J. (2004b). The Effect of Creating External Representations on the Efficiency of Web Search. *Interactive Learning Environments*, *12*(3), 227–250. doi:10.1080/10494820512331383439

Leu, D. J., Coiro, J., Castek, J., Hartman, D. K., Henry, L. A., & Reinking, D. (2008). Research on instruction and assessment in the new literacies of online reading comprehension. In C. Block & S. R. Parris (Eds.), *Comprehension instruction : research-based best practices* (2nd ed.). New York: Guilford Press.

Liang, J.-C., & Tsai, C.-C. (2009). The information commitments toward web information among medical students in Taiwan. *Journal of Educational Technology & Society*, *12*(1), 162–172.

Lim, S. (2009). How and why do college students use Wikipedia? *Journal of the American Society for Information Science and Technology*, *60*(11), 2189–2202. doi:10.1002/asi.21142

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development*, *49*(2), 23–40. doi:10.1007/BF02504926

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. doi:10.1037/0003-066X.57.9.705

Manlove, S., Lazonder, A. W., & Ton de Jong. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition & Learning*, *2*(2/3), 141–155. doi:Article

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Commun. ACM*, *49*(4), 41–46. doi:10.1145/1121949.1121979

Marshall, C. C. (1997). Annotation: From paper books to digital Library. *Proceedings of the second ACM conference on digital libraries* (pp.131-140). Philadelphia, PA: ACM Press. doi:10.1145/263690.263806

Martindale, C. (1981). *Cognition and consciousness*. Dorsey Press.

Mason, L., Boldrin, A., & Ariasi, N. (2009). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607–633. doi:10.1007/s11251-008-9089-y

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, *5*(1), 67–90.

McCrudden, M. T., & Schraw, G. (2007). Relevance and Goal-Focusing in Text Processing. *Educational Psychology Review*, *19*(2), 113–139.

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy*, *4*(3), 331–344.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. doi:10.1002/asi.20672

Miyata, Y., & Norman, D. A. (1986). The control of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design : new perspectives on human-computer interaction* (pp. 265–284). Hillsdale, N.J.: L. Erlbaum Associates.

Moraveji, N., Morris, M. R., Morris, D., Czerwinski, M., & Riche, N. (2011). ClassSearch: Facilitating the development of web search skills through social learning. *Proceddings of the 2011 CHI conference*. Retrieved from hci.stanford.edu/publications/2011/ClassSearch/ClassSearch-CHI2011.pdf

National Research Council. (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology*. Washington, DC: National Academy Press.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. doi:10.1037/0022-0663.97.2.157

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education*, *30*(15), 1977–1999.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020. doi:10.1002/tea.20035

Osman, M. E., & Hannafin, M. J. (1992). Metacognition research and theory: Analysis and implications for instructional design. *Educational Technology Research and Development*, *40*(2), 83–99. doi:10.1007/BF02297053

Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, *91*(2), 230–240. doi:10.1037/0022-0663.91.2.230

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*(3), 423–451.

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What Research Has to Say About Reading Instruction* (pp. 291–309). Newark, DE: International Reading Association.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, *92*(5), 513–555.

Pressley, M., & Lundeberg, M. (2008). An invitation to study professionals reading professional-level texts: A window on exceptionally complex, flexible reading. In K. B. Cartwright (Ed.), *Literacy processes: Cognitive flexibility in learning and teaching* (pp. 165–187). New York: Guilford.

Quintana, C., Meilan Zhang, & Krajcik, J. (2005). A Framework for supporting metacognitive aspects of online inquiry yhrough software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. doi:10.1207/s15326985ep4004_5

Quintana, C., Shin, N., Norris, C., & Soloway, E. (2006). Learner-centered design: Reflections on the past and directions for the future. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 119–134). Cambridge, MA: Cambridge University Press.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A Framework for supporting metacognitive aspects of online inquiry yhrough software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. doi:10.1207/s15326985ep4004_5

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science & Technology*, *53*(2), 145–161.

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science & Technology*, *41*(1), 307.

Rieh, S. Y., & Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 49–71). Cambridge, Mass. : MIT Press, c2008.

Rouet, J. (2006). *The skills of document use: From text comprehension to Web-based learning*. L. Erlbaum Associates.

Rouet, J. F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction*, *21*(2), 205–219.

Rouet, J.-F. (2009). Managing cognitive load during document-based learning. *Learning and Instruction*, *19*(5), 445–450. doi:16/j.learninstruc.2009.02.007

Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85–106.

Rouet, J.-F., Vidal-Abarca, E., Erboul, A. B., & Millogo, V. (2001). Effects of Information Search Tasks on the Comprehension of Instructional Text. *Discourse Processes*, *31*(2), 163–186.

Russell, D. M., Stefik, M. J., Pirolli, P. L., & Card, S. K. (1993). The cost structure of sensemaking (pp. 269–276). Presented at the ACM SIGCHI conference on human factors in computing systems.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science & Technology*, *58*(13), 2126–2144. doi:10.1002/asi.20681

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition∗. *Information Processing & Management*, *26*(6), 755–776. doi:10.1016/0306-4573(90)90050-C

Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate student evaluate web sources. *Reference Services Review*, *26*(3/4), 53–60.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*(1-2), 113–25.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, *15*(1), 27. doi:10.1080/10494820600996972

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. (Ed. . Estes (Ed.), *Handbook of learning and cognitive processes: VI. Linguistic functions in cognitive theory* (Vol. 5, pp. 271–295). Oxford England: Lawrence Erlbaum.

Smith, A. G. (1997). Testing the Surf: Criteria for Evaluating Internet Information Resources. *Public-Access Computer Systems Review*, *8*(3), 1–14.

Stadtler, M., & Bromme, R. (2004). Laypersons searching for medical information on the web: The role of metacognition. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (p. 1638). Mahwah, NJ: Erlbaum.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, *2*(2-3), 191–210.

Stadtler, M., Scharrer, L., & Bromme, R. (2011). How reading goals and rhetorical signals influence recipients' recognition of intertexual conflicts. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Sceicne Society* (pp. 1346–1351). Austin, TX: Cognitive Sciecne Society.

Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities*, *31*(4), 344–64.

Stromoso, H. I., & Braten, I. (2002). Norwegian law students' use of multiple sources while reading expository texts. *Reading Research Quarterly*, *37*(2), 208–227.

Suwa, M., & Tversky, B. (2002). External representations contribute to the dynamic construction of ideas. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), (pp. 341–343). Presented at the LECTURE NOTES IN COMPUTER SCIENCE, Berlin: Springer.

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research (07408188)*, *27*(2), 222–248. doi:10.1016/j.lisr.2005.01.005

Taylor, R. S. (1967). Question-negotiation and information-seeking in libraries, by Robert S. Taylor. Bethlehem, Pa., Center for the Information Sciences, Lehigh University, 1967.

Taylor, R. S. (1968, May). Question-negotiation and information seeking in libraries.

Tsai, C.-C. (2004). Information commitments in Web-based learning environments. *Innovations in Education and Teaching International*, *41*(1), 105–112.

Van der Pol, J., Admiraal, W., & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 339–357.

Van Eemeren, F. H. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. L. Erlbaum Associates.

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning & Instruction*, *1*(4), 337–350.

Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the Web: Students Online in a Sixth-Grade Classroom. *The Journal of the Learning Sciences*, *9*(1), 75–104.

Walraven, A., Brand-gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, *24*(3), 623–648. doi:doi: 10.1016/j.chb.2007.01.030

Walton, D. (2008). A dialogical theory of presumption. *Artificial Intelligence and Law*, *16*, 209–243. doi:10.1007/s10506-008-9063-7

Wang, P., & White, M. D. (1999). A cognitive model of document use during a research project (Study II): Decisions at the reading and citing stages. *Journal of the American Society for Information Science*, *50*(2), 98–114.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301–311. doi:10.1037/0022-0663.91.2.301

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. doi:10.1037/0022-0663.83.1.73

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346. doi:10.1016/S0364-0213(99)80043-3

Winne, P. H. (1983). Training students to process text with adjunct aids. *Instructional Science*, *12*(3), 243–266. doi:10.1007/BF00051747

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives (2nd ed.).* (pp. 153–189). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York: Routledge.

Wolfe, J. (2008). Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 141–164. doi:10.1007/s11412-008-9040-x

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *17*(2), 89–100. doi:10.1111/1469-7610.ep11903728

Wopereis, I., & van Merrienboer, J. J. G. (2011). Evaluating text-based information on the World Wide Web. *Learning and Instruction*, *21*(2), 232–237.

Wu, Y.-T., & Tsai, C.-C. (2005). Information commitments: Evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, *21*(5), 374–385. doi:10.1111/j.1365-2729.2005.00144.x

Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, *58*(1), 181–196.

Zimmerman, B. J., & Tsikalas, K. E. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning? *Educational Psychologist*, *40*(4), 267–271. doi:10.1207/s15326985ep4004_8

CHAPTER 3

THE EFFECTS OF SOURCE REPRESENTATIONS AND GOAL INSTRUCTIONS ON COLLEGE STUDENTS' INFORMATION EVALUATION BEHAVIOR CHANGE [2]

[2] Kim, S. M., & Hannafin, M. J. Submitted to *Journal of the Learning Sciences*

Abstract

This study examined how differing instructional interventions—source representation scaffolds and goal instructions— influence college students' information evaluation behavior (IEB) over time. We conducted a longitudinal, quasi-experimental study with two sections of a large enrollment, introductory general education biology course at a large public university in the southeast United States. Students' self-report measures were collected four times during an academic semester. In order to identify the specific nature of change in IEB score, multilevel analyses were conducted. On average, students' IEB score increased slowly within a baseline group that was given only a checklist and persuasion goals. Both online annotation and balanced reasoning goals yielded improved gains in IEB scores, but there was negative interaction effect when two treatment conditions were combined. The results supported the assumed benefits of online annotation and balanced reasoning goals, yet suggested addressing possible difficulties in using multiple scaffolds. We discussed further implications for research and practice.

Keywords: information evaluation behavior, goal instructions, source representation scaffolds

Introduction

Information evaluation is knowledge and skills necessary but often taken-for-granted in education (Brem, Russell, & Weems, 2001; Iding, Crosby, Auernheimer, & Barbara, 2008; Wiley et al., 2009). Studies emphasized information evaluation during unfamiliar information seeking (Tabatabai & Shore, 2005; Walraven, Brand-gruwel, & Boshuizen, 2008) and richer mental model construction in history (Bråten, Strømsø, & Britt, 2009; Wineburg, 1998), science (Stadtler & Bromme, 2007; Wiley et al., 2009), and law (Stromoso & Braten, 2002). Informaiton evaluation has become increasingly important when students must search, select, and integrate largely unregulated sources of information on the Web to complete academic tasks (Apedoe & Reeves, 2006; Hannafin, Hannafin, & Gabbitas, 2009; Jones, 2001; Metzger, Flanagin, & Zwarun, 2003).

Despite the importance and popularized awareness, information evaluation skills appear not to be readily activated. Even college students often simplify their inquiry tasks, lowering evaluation standards and cognitive engagement (Asher, 2011; Head & Eisenberg, 2009b) and use only superficial evaluation criteria and cues (Bråten et al., 2009; Britt & Aglinskas, 2002). More importantly, despite awareness of the importance of critical evaluation, limited cognitive capacity hinders retaining and connecting source representation during continued information-seeking processes (Eysenbach & Köhler, 2002; Kalyuga, 2009; Rouet, 2009). We need to support learners to employ appropriate information evaluation behavior to address complex information that extends beyond their knowledge threshold.

Information Evaluation Behavior

In this study, we differentiate information evaluation skills from activated behaviors. We examine support for information evaluation behavior in the context of unfamiliar learning tasks.

In order to do so, it is necessary to understand related mental tasks and conditions. Information evaluation encompasses a series of mental tasks related to representation and control of one's *knowledge sources* (e.g., "is this source credible enough to support claim X?"); this is largely metacognitive in nature (see for associated discussion, Hofer, 2004; Kuhn, 2000; Mason, Boldrin, & Ariasi, 2010). Accordingly, information evaluation behavior involves a) eliciting standards in light of one's task and goals, b) representing and comparing cues (i.e., sources properties) against standards, and c) interpreting the difference to match an appropriate action (i.e., re-reading, marking source distinctions, and seeking additional sources) (c.f., Pressley, 2002; Rouet, 2006; Schraw, 1998; Winne & Nesbit, 2009). Though many describe information evaluation as a skill set needed at one stage of sequential information-seeking processes (see for a review, Lazonder & Rouet, 2008), information evaluation behavior involves iterative representation and control processes with four changing foci as one's cognitive state and information needs change over time (c.f., Saracevic, 2007; Schamber et al., 1990). Internal conditions such as task perceptions and working memory capacity mediate these cognitively demanding processes (c.f., Tanni & Sormunen, 2008; Winne & Hadwin, 1998).

Predictive evaluation. Given a task (e.g., "write a newspaper column that explains which molecules build up in muscle cells causing them to burn during exercise"), people define their goals and elicit previous experiences about where and how to find specific sources, thereby setting expectations for needed sources even before seeking information (Hogarth, 1980; Hughes, Wareham, & Joshi, 2010; Rieh & Hilligoss, 2008; Wu & Tsai, 2005). For example, students often define inquiry tasks as findings 'facts.' Accordingly, students seek 'safe sources' and favor only instructor-selected sources to conduct a research assignment (c.f., Alison & Michael, 2009; Bråten, Strømsø, & Salmerón, 2011). As such, predictive evaluation can either limit or enhance information seeking scope and strategies (c.f., Hughes, Wareham, & Joshi, 2010; Wu & Tsai, 2005). It is notable that task perceptions orient initial information-evaluation behavior. For example, task value (e.g., interest, importance, and utility) is an enabling factor of monitoring efforts and cognitive engagement (c.f., Lan, 2005; Wigfield & Eccles, 2000). If students consider a task as important, monitoring tends to increase. Perceived goals serve as standards that guide monitoring and regulation of learning processes (Jiang & Elen, 2011; Winne & Hadwin, 1998). Some researchers further manipulated students' perceived goals using adjunct questions (Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009), directions (Nussbaum & Kardash, 2005), or prompts (Belland, Glazewski, & Richardson, 2010) that help with goal articulation.

Intrinsic evaluation. If the task is unfamiliar, information evaluation at the initial stage does not necessarily account for logical coherence or strength of sources. Rather, by focusing on the intrinsic, or innate merit of sources such as currency, support, authority, and bias (Rieh & Hilligoss, 2008; Wineburg, 1998), evaluation helps to familiarize with terminologies and narrow the focus readily (e.g., "what is known about burning sensation during exercise?") (Russell, Stefik, Pirolli, & Card, 1993). Intrinsic evaluation may have little strength in terms of fit for a specific task context, but serves as a minimum qualification to locate sources (Tabatabai & Shore, 2005; Walraven et al., 2008). It therefore reduces extraneous cognitive load caused by trial and error strategies to locate relevant resources (c.f., Kalyuga, 2011). Such intrinsic evaluation has been the strong focus of evaluation training in education and library studies (Hjørland, 2012; Meola, 2004).

Extrinsic evaluation. With increased familiarity with a given task, task goals and information needs become clearer. Extrinsic evaluation helps to achieve goals by weighing relative strength and utility of each source (Hjørland, 2012; Meola, 2004). Text comprehension studies report such extrinsic evaluation helps to integrate conflicting sources and create a coherent mental model (Bråten et al., 2009; Wiley et al., 2009). For instance, students may find conflicting claims (e.g., "lactic acid is not a waste product that causes burning sensation during exercise"). By marking and comparing source distinctions, students can integrate different claims (e.g., "there is however *another* explanation according to *Dr. so and so*") and further proceed to solve the inconsistency (e.g., "this alternative source provides *stronger support*"). However, such multiple representations intrinsically demand high cognitive load (c.f., Kalyuga, 2009; Rouet, 2009). Thus, in Eysenbach and Köhler (2002), adults who searched for health information rarely activated evaluation strategies but were able to articulate their strategies when asked to do so. College students, though they consider information quality important (Head & Eisenberg, 2010), experienced difficulties when combining multiple sources to reorganize incompatible information. Carefully designed scaffolds may augment cognitive capacity by segmenting, sequencing, and finally automating required mental activities (c.f., Scheiter, Gerjets, & Schuh, 2010).

Reflective evaluation. Following a series of information evaluations, summative reflection on the range of collected sources may prove beneficial. Metacognition researchers suggest such reflection may enhance short as well as long-term task performance (Pressley, 2002; Schraw, 1998). Summative reflection especially supports reorganization of conflicting information (c.f., Russell, Stefik, Pirolli, & Card, 1993) to refine the final product. Reflection may also trigger the individual to reconsider the relevance of existing knowledge of cues, criteria, or sources. Such metacognitive knowledge influences future information evaluation behavior (Bowler, 2010; Rieh & Hilligoss, 2008; Schraw, 1998).

Source Representation Scaffolds

Metacognitive studies emphasize that training in monitoring one's performance alone, may establish a self-sustaining metacognitive development mechanism and thus benefit learning (Borkowski, Carr, & Pressley, 1987; Pressley, Borkowski, & Schneider, 1987; Schneider & Pressley, 1997). Information and library researchers have documented efforts to help students explicitly represent source quality properties (see for the review, Hjørland, 2012; Meola, 2004) to better manage information seeking. Checklists that provide lists of evaluation criteria are among the most common aids. Checklists typically incorporate acronyms—for example, CRAP representing Currency, Reliability, Authority, and Purpose ratings—to facilitate the internalization of evaluation criteria. Though simple and efficient, checklists do not typically provide explicit cues and heuristics for a given task context (e.g., argumentation), limit holistic representation across multiple sources, and do not address cognitive hurdles (e.g., simplifying inquiries or reducing cognitive load) student face, sometimes promoting simple mechanical analyses (Harris, 2007; Meola, 2004; Metzger, 2007). Computer scaffolds when embedded within a specific task, on the other hand, help learners to represent and interpret task-related cues while minimizing extraneous cognitive load (Gama, 2004; Roll, Aleven, McLaren, & Koedinger, 2007; Winne & Nesbit, 2009). Online annotations, for example, can represent source features by providing structure or directions within a task (e.g., argumentation), may serve to model expert use of evaluation cues in a high contextual manner (c.f., Cho & Jonassen, 2002; Oh & Jonassen, 2007; Reiser, 2004). Annotation tools may further assist in documenting information externally, thereby allocating working memory effectively during information seeking and evaluation (c.f., Belland, Glazewski, & Richardson, 2008; Lee, 2004; Miyata & Norman, 1986; Zhang & Quintana, 2012).

Goal Instructions

Information evaluation processes interact with individually perceived task goals, which guide decisions whether to seek local coherence within a single source (i.e., fact-finding) or global coherence across sources (i.e., convergence and elaboration) thereby specifying evaluation standards and allocating cognitive efforts (Kuiper, Volman, & Terwel, 2005; Stadtler, Scharrer, & Bromme, 2011). Researchers suggest that simply changing the specificity of written directions for a given task can influence students' task-related behavior (Ferretti, MacArthur, & Dowdy, 2000; Nussbaum & Kardash, 2005; Page-Voth & Graham, 1999). For instance, in Nussbaum and Kardash (2005), students were asked to write an essay about the harmful effects of TV violence and given three different goals: baseline ("explain your claim"), reasoning ("provide as many reasons as possible"), and rebuttal ("provide the reason why your counterarguments are wrong"). Results indicated that students wrote more counter-arguments and rebuttals given rebuttal-goal instructions than the other conditions. The authors also cautioned against using a general persuasion goal ("convince or persuade your friend about your claim") since this goal may induce students' efforts to over-convince the audience ignoring counter-arguments. In previous studies, tasks (e.g., reading, argumentation) rarely involved the free search for information. The implications, however, appear especially relevant to information evaluation during online inquiries. Unless asked to intentionally cross-validate different sources of information, students are likely to process incomplete or conflicting information mistakenly. A more critical stance ('comparing different sides') may promote evaluation across sources and verification of source validity in comparison with other sources.

Overview of the Study

This study examined how differing instructional interventions influence information evaluation behavior over time. A two-factorial, quasi-experimental study with repeated measures was designed (Kirk, 1994) varying source representation scaffolds (factor A) and goal instructions (factor B). Students given a checklist and persuasion goals ("persuade or convince the public of your claim") served as a baseline group. For treatment, we employed an online annotation tool or/ and balanced reasoning goals ("critically compare various claims" or "provide the reason why your counterarguments are wrong"). We collected self-report measures of Information Evaluation Behavior (IEB) four times throughout an academic semester. We also collected perceived task values and working memory capacity measures to explain individual variability in IEB scores. Figure 3.1 summarizes the study. Research questions were:

1. How does information evaluation behavior change over time?
2. How do varied interventions affect information evaluation behavior?

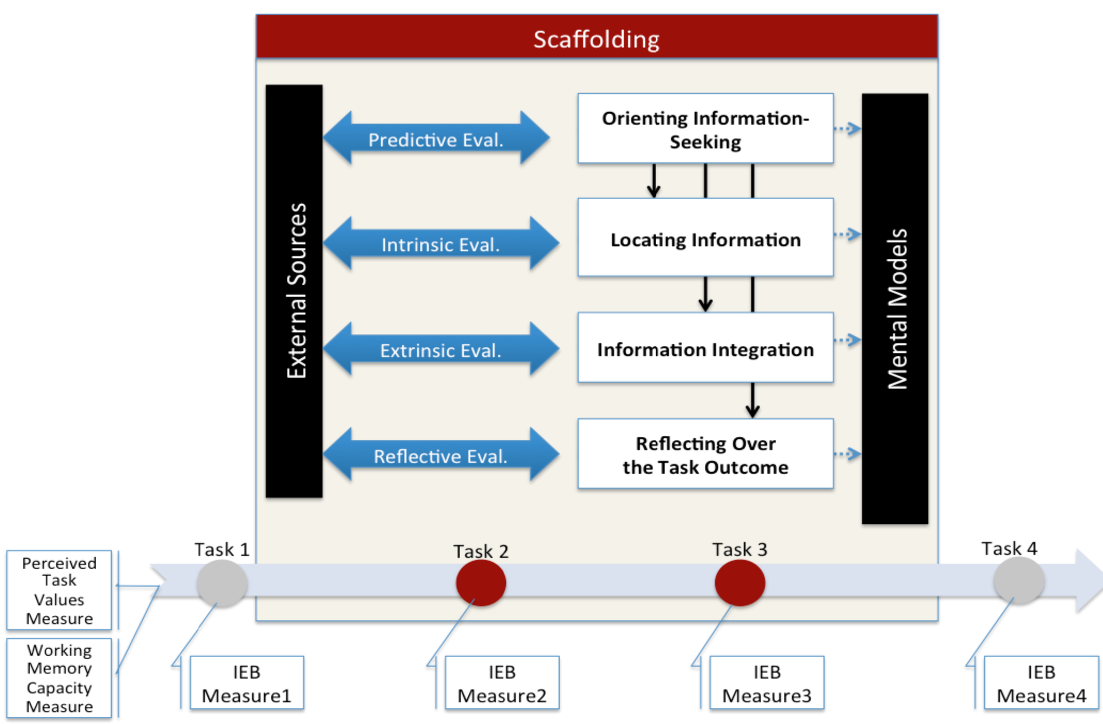

Figure 3.1. Study overview

Research Method

Settings and Participants

The study was conducted in Spring 2012 using two intact sections of a large-enrollment introductory biology course at a large public university in the southeast United States. The original class size was 606. The current study included 384 students (68%) who voluntarily responded to the all items comprising four IEB measures. Students were given extra credits upon completion of each measure. According to a demographics survey (see Appendix A) implemented in the beginning of the semester, the participants included 75% female, 74% freshmen (average age 19), and 89% non-science majors. Before this class, they completed an average of one college level science course. Ninety-one percent of students scored 50% or lower on a prior knowledge test of important concepts covered in the class (refer to Appendix B for the test items); their mean score was 6.46 of a maximum possible score 18. This ensured that students did not already possess the target knowledge to be learned from inquiry tasks. The original and final samples did not vary significantly in terms of demographics as well as prior knowledge, perceived task value, and working memory capacity test scores.

Materials and Procedures

We invited the entire student enrollment from two sections to participate in the study at the beginning of the semester. The same instructor taught both sections using the same instructional materials and procedures. The study was conducted in naturalistic classroom settings. All students were exposed to the checklist use as part of classroom interactions. Each section was then assigned to either checklist only (control) or online annotation tool + checklist condition (treatment) (factor A). Project groups comprising 5-6 students were randomly assigned either to a persuasion goal condition (control) or balanced reasoning goal condition (treatment) (factor B). Given the widespread use of checklists and persuasion goals in science classes, those interventions supported the baseline group. Table 3.1 summarizes the design scheme and number of students assigned to each condition.

Table 3.1

A Quasi-Experimental, Two-Factorial Design Scheme with Repeated Measures

Source Representation Scaffolds (Factor A)	Goal Instructions (Factor B)	N	IEB Measures (Measurement time points/week)			
			1st(2)	2nd(4)	3rd(6-14)	4th(16)
Checklist only (C)	Persuasion (C)	85				
	Balanced Reasoning (T)	85				
Annotation plus checklist (T)	Persuasion (C)	108	-	1st INT	2nd INT	-
	Balanced Reasoning (T)	106				
Total		384				

Note. C and T represent control and treatment conditions respectively. Interventions (INT) were provided at the 2nd and 3rd measures. Students completed the 1st, 2nd, and 4th measures at week 2, 4, and 16 respectively. The 3rd measurement time varied across chosen inquiry tasks (week 6 to week14).

Inquiry Tasks. The course emphasized an inquiry curriculum, where students used Web resources to examine different aspects of biological phenomena to dispute prevalent misconceptions. For example, students were asked to write a newspaper health column that explains which molecules build up in muscle cells causing them to burn during exercise (refer to Appendix C for inquiry topics)**.** During week 2, all students conducted the first inquiry task without interventions; the task was used to document entry information evaluation skills in. Students were then trained in using a checklist or/and the online annotation tool (see Appendix D and E for training materials). Interventions were provided along with the second and third tasks and removed during the final task. Immediately after each task, students rated their individual information evaluation behavior using the IEB scale online. The first and fourth tasks were individual assignments while the second and third tasks were group projects. Independent of task types, the unit of analysis was individual student change over time. Since the third task was chosen among six topics bi-weekly distributed across the semester, the time interval between tasks was not identical across students. For example, student 1 conducted tasks at week 2, week 4, week 6, and week 16. Student 2 conducted tasks at week 2, week4, week 14, and week 16. Discontinuity in intervention implementation and differing schedules for the 3rd task resulted in variations in cumulated intervention effects over time for individual students.

Source representation scaffolds. The written checklist (control) provided a list of evaluation criteria students could use to examine information sources. The criteria included CRAP—currency, reliability, author, and bias in a low contextual manner (Britt & Aglinskas, 2002; Wiley et al., 2009; refer to Appendix F for the checklist details). The instructor continued to emphasize CRAP testing during class sessions to automatize information evaluation. All students, individually and in groups, submitted their source evaluation rationale after each inquiry task. In contrast, the online annotation tool—*Showing Evidence*[3]—provided a visual framework for identifying, judging, and linking evidence that supports or weakens claims within argumentation structure (see Figure 3.2). Students in the treatment used the annotation tool to pool documents from the Internet search and integrate them to their argumentation. (a) Students either had their own claim from the beginning or created a temporary claim from scanning collected documents. (b) Students registered each document in the evidence bin to (c) record and judge source properties (e.g., author, date, and abstract) using a 5-point rating system for intrinsic quality (e.g., It is basically about OOO and is from a qualified source because...”). (d) Students then visually linked each source to the claim and judged how strongly identified evidence supports or weakens the claim using a 5-point rating system. Students also recorded

[3] The authors acknowledge Intel® Innovation in Education for permission to use the tool in this research (refer to Appendix G for the permission)

their reasoning (e.g., “I believe this strongly weakens my claim since…”), which served as a warrant statement for their argumentation. (e) Students finally evaluated the entire information collection (e,g., “is my collection integrated into a consistent whole?”) to rate the strength of their claim and to (f) draw a final conclusion. Individual annotation results were shared to create a group annotation, thereby creating a final group paper.

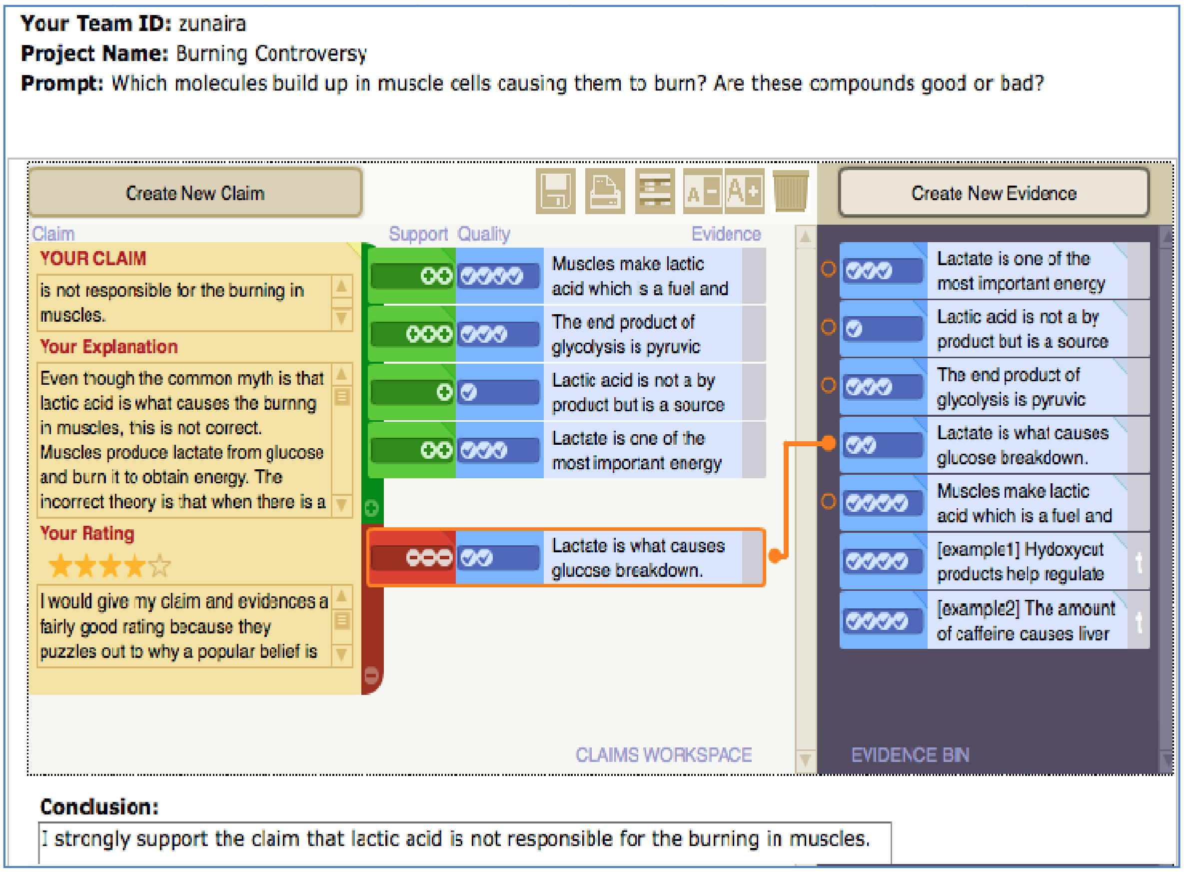

Figure 3.2. Showing Evidence annotation tool screen shot. Students completed six steps: (a) Creating claim; (b) Creating evidence; (c) Rating evidence quality (intrinsic evaluation); (d) Rating evidence strength (extrinsic evaluation); (e) Rating the strength of claim (reflective evaluation); and (f) Drawing a final conclusion.

We expected that students who used the annotation tool would develop more elaborated information evaluations since it integrated evaluation dimensions within an argumentation task, decomposed the process, externalized the advanced cues for extrinsic and reflective as well as intrinsic evaluation, and minimized cognitive load.

Goal instructions. To address concerns regarding predictive evaluation, written goal statements were embedded in the second and third project instructions (Refer to Appendix H for specific statements). For the first and fourth tasks, no explicit cues guided students regarding the task goals. The persuasion goal (control) asked students to write a persuasive letter or convince their peers to accept their claim. The balanced reasoning goal (treatment) asked students to support an informed decision of the audience while critically examining opposing arguments and multiple points of view. We expected that balanced reasoning instructions would induce higher evaluation standards (e.g., "is there any significant counter-evidence left unexplored?") and thus more elaborated information evaluation behavior.

Outcome Measures

Instrument development. A 24-item self-report instrument (Information Evaluation Behavior Scale: IEB; refer to Appendix I) was developed based on the theoretical framework of information evaluation behavior, which consists of 4 sub-dimensions—(a) predictive evaluation (i.e., judging where and how to find information) (Hogarth, 1980; Rieh & Hilligoss, 2008; Rieh, 2002), (b) intrinsic evaluation (Brem et al., 2001; Wiley et al., 2009), (c) extrinsic evaluation (Mason, Boldrin, & Ariasi, 2009; Meola, 2004; Rieh & Hilligoss, 2008), and (d) reflective evaluation (Rieh & Hilligoss, 2008; Russell et al., 1993). Each item was designed to exclusively represent one of 4 sub-dimensions, which assumed to represent one higher-order construct. Ratings for each item were made on a 100-mm, bi-polar scale with the right end indicating the statement was false and the right indicating the statement was true. Continuous scales were used instead of Likert-scales to increase response variation (Schraw & Dennison, 1994). The possible score for each item ranged from 0 to 100.

Two faculty members in information sciences and science education examined the face validity from information seeking and science inquiry perspectives. Five college students participated in a field test to check clarity in wording. A preliminary study was then conducted with 304 undergraduate students recruited from an introductory Biology class in Fall 2011 to confirm the validity of assumed measurement structure and reliability. As a result, four items were dropped due to lack of clarity and redundancy. Four items were dropped due to low item-to-total correlation (<.3). Final 16 items yielded a good internal consistency with the alpha level .91. As we hypothesized, items loaded on four factors with good coefficients (.56 to .76). Four factors were again strongly correlated (.86 to .96) and loaded on a 2nd order factor with factor loadings .90 to .97 implying items ultimately represent one construct. Fit indices—RMSEA (<.09), SRMR (<.09), NNFI (>0.95), IFI (>.95), and CFI (>.95)—showed this structure fits the actual data structure well. Table 3.2 summarizes factors, definitions, and item examples.

Table 3.2

Information Evaluation Behavior (IEB) Scale Summary

2nd order Factor	1st order Factor	Definition	# of Items	Item Example	Reference
Information Evaluation Behavior	Predictive Evaluation	Expectation set about needed sources	3	"I asked myself whether different perspectives might exist in relation to the given topic.”	Hogarth, 1980; Rieh, 2002
	Intrinsic Evaluation	Evaluation of innate merit of sources	5	“I asked whether the author had proper authority.”	Brem et al., 2001; Wiley et al., 2009
	Extrinsic Evaluation	Evaluation of relative strength and utility of sources	5	“I weighed different sources of information to strengthen my claim.”	Hjørland, 2012; Meola, 2004
	Reflective Evaluation	Summative evaluation of source collections	3	“I periodically paused to draw an overall picture of multiple documents that I had selected.”	Rieh & Hilligoss, 2008; Russel, Stefik, Pirolli, and Card, 1993

Information Evaluation Behavior (IEB). IEB measures were collected online four times, each right after an inquiry task retrospectively reflecting their individual information behavior during the task. The scale yielded a good internal consistency in the current study. The average alpha level over four measurements was .93. Due to unbalanced number of items across dimensions, an index of information evaluation behavior was created summing average scores of sub-dimension items to reflect each dimension equally. The possible sum scores accordingly ranged 0 to 400 with higher values indicating better evaluation behaviors.

Level 2 Predictors

Two instruments, designed to measure individual differences in perceived task value and working memory capacity, were administered online at the beginning of the semester. These individual predictors were assumed to explain variability in the initial status and growth slope.

Perceived Task Value (PTV). We measured perceived task value in science using three Science Motivation sub-scales (Glynn, Taasoobshirazi, & Brickman, 2009; refer to Appendix J) focusing on perceived interest, importance, and utility (c.f., Wigfield & Eccles, 2000). Students rated a 15-item, 5-point Likert scale to assess the degree to which each statement characterizes their perception (e.g., “I like science that challenges me”). The alpha reliability coefficient was .91. An index score was created by summing individual item scores and then centered on the mean (50.66). Accordingly, possible scores ranged from -45.66 to 24.34 after centering. We expected higher scores would predict higher evaluation behavior.

Working Memory Capacity (WMC). Working memory capacity, the ability to process incoming information, was considered as a possible factor to facilitate multiple source representations. The operation span task (Turner & Engle, 1989; refer to Appendix K) was used to measure the degree to which students can retain words while solving simple math problems (for example, does [2/2] + 4 = 6?). WMC contains 15 sets of operation tasks, constituting 60 brief math problems and words. In order to reduce ceiling effects, a 4.5 second time limit was set for each math problem. The alpha reliability coefficient was .90. The score averaged proportion of correctly recalled elements within each set (Conway et al., 2005). The possible score ranged from 0 to 1, and the mean score was .76. Considering scale differences between WMC and other measures, the score was multiplied by 100 and then centered on the adjusted mean (76). The possible score ranged from -76 to 24 after centering.

Data Analysis

Repeated information evaluation behavior measures were nested within individuals. We conducted a longitudinal analysis with two-level mixed models with random coefficients. While the repeated measures ANOVA can determine possible intervention-related change, the model does not include time variables or time-varying predictors and thus cannot explain the specific nature of change. Also, the repeated measures ANOVA assumes balanced data and common covariance between measurement occasions. These assumptions restrict flexible modeling of unbalanced and unstructured data.

As an alternative, we partitioned overall variance into 2 levels to explain within-person variation over time (Level 1) and between-person variation (Level 2). We sequentially built and tested a series of growth models by adding time, time-varying predictors, and time-invariant predictors in order and conducted a deviance test to determine whether additional predictors might yield a better model fit. Full information ML (FIML) was used to compute deviance statistics. In contrast, parameters were estimated using residual ML (REML) since REML estimates unbiased variance. The Kenward-Roger method approximated denominator degrees of freedom, thus reproducing exact F tests for unbalanced designs. Final inferences were generated using a parsimonious model with the best fit for the given data. Alpha level was set to .05 for all tests.

Results and Discussion

The preliminary analysis revealed correlations among outcome measures typical of longitudinal patterns (Table 3.3). Correlations among measures were positive at different points, but decreased as the interval between measures increased. This pattern supported the use of multi-level modeling for correlated yet largely unstructured observations, which do not follow the common covariance assumption of repeated measures ANOVA.

Table 3.3

Outcome Measures Correlation Matrix

	IEB1	IEB2	IEB3	IEB4
IEB1 (1st Wave)	1 (4,483)	-	-	-
IEB2 (2nd Wave)	.74** (3,055)	1 (3,616)	-	-
IEB3 (3rd Wave)	.52** (1,792)	.71** (2,193)	1 (2,945)	-
IEB4 (4th Wave)	.50** (1,994)	.66** (2,432)	.76** (2,300)	1 (3,684)

Note. ( ) indicates covariance. **correlation significant at alpha=0.01 level (two-tailed).

Table 3.4

Descriptive Statistics of Level 1 and Level 2 Continuous Variables (N=384)

Variables	Item N	Mean	S.D.	Skewness	Kurtosis	Cronbach's α
IEB1 (1st Wave)		267.34	66.96	-.48	-.10	.912
IEB2 (2nd Wave)	16	306.14	60.13	-1.00	1.58	.923
IEB3 (3rd Wave)		315.69	54.26	-.66	.13	.923
IEB4 (4th Wave)		314.01	66.69	-1.19	2.53	.932
PTV_Raw	15	50.66	9.20	-.12	-.09	.91
WMC_Raw	15	.76	.15	-1.60	3.21	.90

Note. IEB=sum of average dimension scores (item *N*=16). PTV_Raw and WMC_Raw indicate perceived task value and working memory capacity raw scores before mean centering.

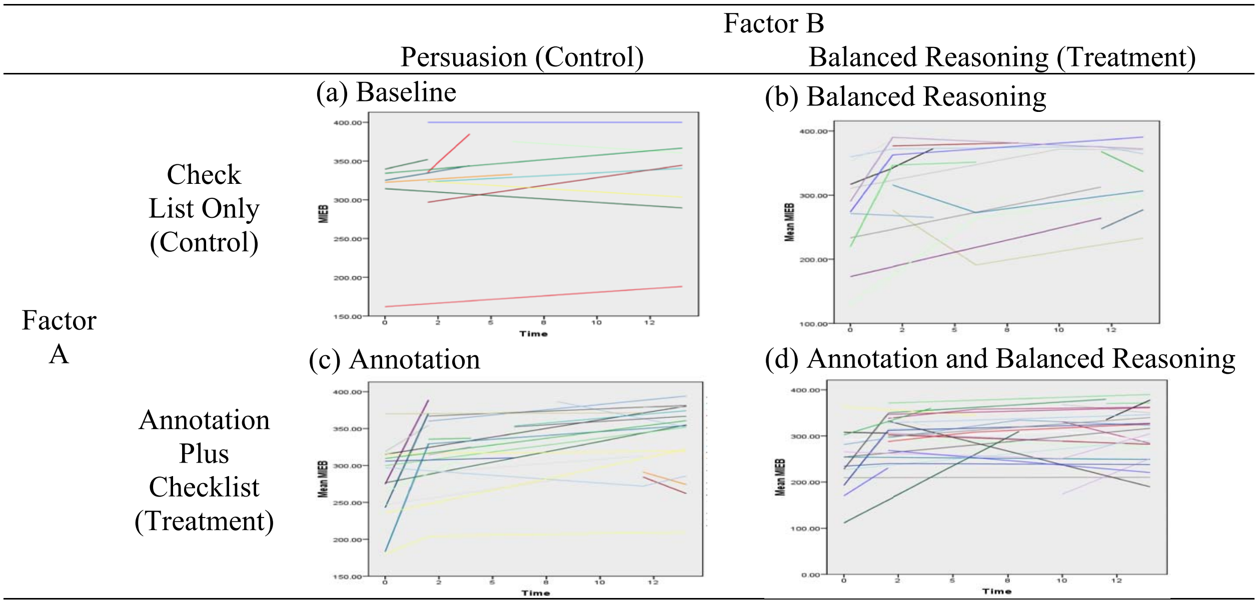

Figure 3.3. Growth trajectories for each student classified by instructional interventions. The X and Y-axes indicate time (TIME =0, 2, 4, 6, 8, 10, 12, 14weeks) and IEB scores (min=0, max=400) respectively.

Mean scores increased over the semester from 267.34 to 314.01 as shown in Table 3.4. Individual students' growth trajectories in Figure 3.2 exhibited similar upward trends independent of intervention. Baseline students given a checklist and persuasion goals exhibited relatively simple and gradual linear growth while the growth of the remaining students fluctuated with pronounced elevations at the 2nd measurement. Scores stabilized or decreased slightly at the final measurement. We assumed differentials appeared due to differing interventions introduced at the 2nd and 3rd measurement occasions. Score differentials persisted over time even after the intervention was removed (c.f., McCoach & Kaniskan, 2010). However, we assumed such persistency would vary across time considering discontinuity in interventions and differing schedules for the 3rd task. In addition, Figure 3.2 demonstrated individual variability in the initial status and growth rate, which was in line with the intraclass correlation indicating 57% of between-person variation over the overall variance. Level 2 predictors—perceived task values and working memory capacity scores—exhibited significant correlations with outcome measures (.21 and .17 respectively; $p<.05$).

Table 3.5

Summary of Variables

Variable Label	Description	Coding / Score Range
Outcome Variable		
Y_{it}	Student i's Information Evaluation Behavior score at measurement occasion t	Possible score range: 0 – 400
Level 1 Predictors		
$(Time)_{it}$	Week of measurement occasion t, centered on week2	TIME: Week 0, 2, 4, 6, 8, 10, 12, 14
$(SOURCE)_{it}$	Exposure to source scaffolds; time-varying	Online annotation: 0, 1, 2, 2; Checklist: 0, 0, 0, 0
$(GOAL)_{it}$	Exposure to goal instructions; time-varying	Balanced Goal: 0, 1, 2, 2; Persuasion Goal: 0, 0, 0, 0
$(SOURCE)_{it} \times (GOAL)_{it}$	SOURCE × GOAL interaction term	
$(SOURCE)_{it} \times (Time)_{it}$	SOURCE × Time interaction term	
$(GOAL)_{it} \times (Time)_{it}$	GOAL × Time interaction term	
Level 2 Predictors		
$(PTV)_i$	Perceived Task Value; mean centered (50.66)	Original score range: 5 to 75 Centered score range: -45.66 to 24.34
$(WMC)_i$	Working Memory Capacity; mean centered (.75)	Original score range: 0 to 1 Centered score range: -.75 to .25

Based on initial data inspection, we applied a piecewise, two-level growth model to append separate intervention-associated slopes to the normal growth slope and then included level 2 predictors. While a quadratic model could capture discontinuity in the growth trajectories, use of separate slopes was intended to improve flexibility and interpretability (c.f., Chou, Yang, Pentz, & Hser, 2004; McCoach & Kaniskan, 2010). Table 3.5 summarizes variables in the full model. At level 1, $(Time)_{it}$, a continuous variable, represented a weekly time frame. Intervention exposure, $(SOURCE)_{it}$ and $(GOAL)_{it}$ were coded 0, 1, 2, 2 for treatment and 0, 0, 0, 0 for control at four consecutive measurement occasions. Since we had four repeated measures, we set only the intercept (initial status) and $(Time)_{it}$ parameter (weekly growth rate) as random across individuals. Thus, intervention exposure effects and their interaction were set to persist over time, consistent with observations, yet with same rates. To make up for the fixation, we included intervention-time interaction terms, $(SOURCE)_{it} \times (Time)_{it}$ and $(GOAL)_{it} \times (Time)_{it}$. For level 2, we included perceived task value $(PTV)_i$ and working memory capacity scores $(WMC)_i$ for variability in initial status and growth slope.

As summarized in Table 3.6, significant IEB predictors were included in the full model. The deviance statistic between the full and reduced models was not significant ($\chi^2(2) = 2$, p>.05) meaning the final reduced model is a parsimonious model with the best fit. The final model therefore included nine fixed-effects and four variance components.

The average initial status (β_{00}) was 274.53 controlling for other effects. In other words, students with average PTV and WMC scores reached 68% level of the possible 400 on average given no intervention. The model yielded positive individual differentials in initial status associated with perceived task value (β_{01}= 1.14) and working memory capacity (β_{02}= .56). Therefore, every one-unit increase above the mean perceived task value and working memory capacity scores resulted in 1.14 and .56 score increase in the initial status respectively.

The average weekly growth rate (β_{10}) was 1.56 controlling for other effects, meaning IEB scores increased by average 1.56 weekly for students with average PTV and WMC scores given the baseline condition. The growth rate corresponded to .4% of the possible 400. There was no significant impact of PTV on individual growth rates (β_{11}). However, there was a negative differential in the weekly growth rate associated to WMC (β_{12} = -.04). Therefore, one unit increase in the individual WMC level above the mean decreased the growth rate by .04.

Table 3.6

The Full and Final Models of Information Evaluation Behavior Change

Parameter			Full Model	Final Model
			Estimate (SE)	Estimate (SE)
Fixed Effects				
Initial Status, π_{0i}	Intercept	β_{00}	274. 04 (3.72) ***	274.78 (3.67) ***
	Slope	β_{01}(PTV)	1.18 (.38)**	1.14 (.33)***
		β_{02}(WMC)	.56 (.23)*	.56 (.22) *
Growth Rate, π_{1i} (Time)	Intercept	β_{10}	1.90(.45) ***	1.56(.39) ***
	Slope	β_{11}(PTV)	-.005 (.03)	-
		β_{12}(WMC)	-.04(.02) *	-.04 (.02) *
SOURCE Effect, π_{2i}	Intercept	β_{20}	16.32(3.70) ***	13.30(3.03) ***
GOAL Effect, π_{3i}	Intercept	β_{30}	27.52(4.05) ***	30.12(3. 60) ***
SOURCE × GOAL, π_{4i}	Intercept	β_{40}	-9.88 (2.29)***	-10.74(2.21)*
SOURCE × Time , π_{5i}	Intercept	β_{50}	-.47(.32)	-
GOAL × Time, π_{6i}	Intercept	β_{60}	-1.01(.34) **	-1.11(.33) **
Variance Components				
Level-1	VAR(e_{it})	σe2	911.18(64.90)***	914.49(65.13)***
Level-2	VAR(Υ_{0i})	τ_{00}	2,964.96(306.64)***	2942.22(304.38)***
	VAR(Υ_{1i})	τ_{11}	8.71(1.67)***	8.44(1.63)***
	Covariance (Υ_{0i}, Υ_{1i})	τ_{01}	-67.68(18.16)***	-63.77(17.74)***
Fit Statistics	-2 LL		10,571.20	10573.20
	# of parameters		15	13
	Deviance (df)			2(2)
Pseudo R2	In level 1 variation			44.78
	In level 2 variation (initial status)			6.39
	In level 2 variation (weekly growth rate)			11.57

Note. Parameters were estimated using REML. *p< .05, **p<.01, ***p<.001.

Differing intervention conditions changed students' weekly growth rate significantly. Given online annotation at the 2nd measurement occasion, for example, students' growth rate increased by 13.30 (β_{20}). Balanced goals increased the growth rate by 30.12 (β_{30}). However, there existed a negative interaction effect (β_{40}=-10.74), which countervailed 26% of the possible sum of treatment effects at the second measurement occasion. Also, the final model included a negative goals-time interaction (β_{60} = -1.11), which indicated a significant decay of balanced goal effects over time. On the other hand, there was no significant online annotation-time interaction (β_{50}).

In terms of variance components, within-person variation (σe2) was = 914.49. Between-person variation in the initial status (τ_{00}) and weekly growth rate (τ_{11}) was 2,942.22 and 8.44 respectively. Level-1 model explained 44.8% of within-person variation while level-2 model explained 17.94 % of between-person variation. The results indicate there are still statistically significant variability unexplained. Interestingly, there was a negative association between τ_{00} and τ_{11} (τ_{01}= -63.77), which means students with higher IEB scores at initial status had less sharp growth slopes.

General Discussion

This study examined how differing instructional interventions—source representation scaffolds and goal instructions—influenced college students' information evaluation behavior over time. A longitudinal quasi-experimental study was conducted with multilevel models analysis. In this section, we interpret the final model to address our research questions.

Question 1: How does information evaluation behavior change over time?

Initial Status and Growth Rate. Overall, information evaluation behavior (IEB) followed a linear, but gradual growth pattern. The average initial IEB score for students was only 68% of the possible 400. The average weekly growth rate was still low in the baseline condition. The checklist and persuasion goals contributed to weekly increases by only .4 %. The results supported a need for additional support to help students employ appropriate information evaluation behavior.

Individual Differences and Information Evaluation. To understand the nature of change, we examined between-person variation. As detected from individual growth trajectories, students demonstrated different initial IEB scores and weekly growth rates. Variation in the initial status was much larger than the weekly growth rate. This distinction indicates that initial status largely mediated individual performance over time. The growth rate varied but was relatively stable across students. However, a negative association was detected between the initial status and growth rate. Students with lower initial IEB scores therefore demonstrated sharper growth slopes suggesting lower performance students benefited from instructional scaffolding (c.f., Belland, Glazewski, & Richardson, 2010).

Both perceived task value (PTV) and working memory capacity (WMC) were significant predictors of variation in the initial status. Increases above the mean PTV score and WMC score yielded corresponding increases in initial IEB scores when controlling for other effects. The results are in line with metacognitive theories which emphasize the influences of task perceptions and cognitive capacity on both monitoring and cognitively demanding task performance (Lan, 2005; Winne & Hadwin, 1998). Scaffolding design, thus, requires deliberate consideration of student interpretation and mental activities—especially related to links among task perception, goal setting, and cognitive load.

While the PTV on IEB score was consistent over time, students with higher WMC scores did not improve as much as students with lower WMC scores. The results are in line with the negative association found between the initial status and growth rate. Our findings may be analogous to expertise reversal effects that occur when redundant instruction to knowledgeable or competent students interferes with learning (Kalyuga & Renkl, 2010; Kalyuga, 2007). The baseline condition may have offered functional redundancy, countervailing benefits that large working memory capacity can provide. For example, checklists easily automate evaluation tasks even though they mainly support simple, intrinsic evaluation (c.f., Hjørland, 2012; Meola, 2004).

Question 2: How do varied interventions affect information evaluation behavior?

The analysis of weekly growth patterns indicated the baseline condition contributed to modest increases in IEB scores, which supported our attempt to examine scaffolding approaches in conjunction with mental tasks. Figure 3.4 illustrates predicted mean score change across intervention conditions and schedules.

Source Representation Scaffolds and Goal Instructions. Overall, treatment conditions elevated IEB scores over time. Online annotation alone yielded a statistically significant differential growth slope (β_{20}= 13.27). Since students were exposed twice to online annotation tools and we set the impact to persist, annotation tool use alone resulted in 13.24 and 26.54 score add-up to the baseline growth lines at the second and third measurement occasion respectively. The differential 26.54 persisted after the intervention removal at the final wave resulting in piecewise growth patterns with varying growth rates over time. Similarly, a significant differential growth slope was associated with balanced reasoning goals (β_{30}= 29.56). The results provided support for the benefits of online annotation and balanced reasoning goals. In addition, a large differential rate in balanced reasoning goals supported the efficiency of this simple intervention (c.f., Nussbaum & Kardash, 2005; Nussbaum, 2005). Students displayed even larger gains than those in the annotation condition.

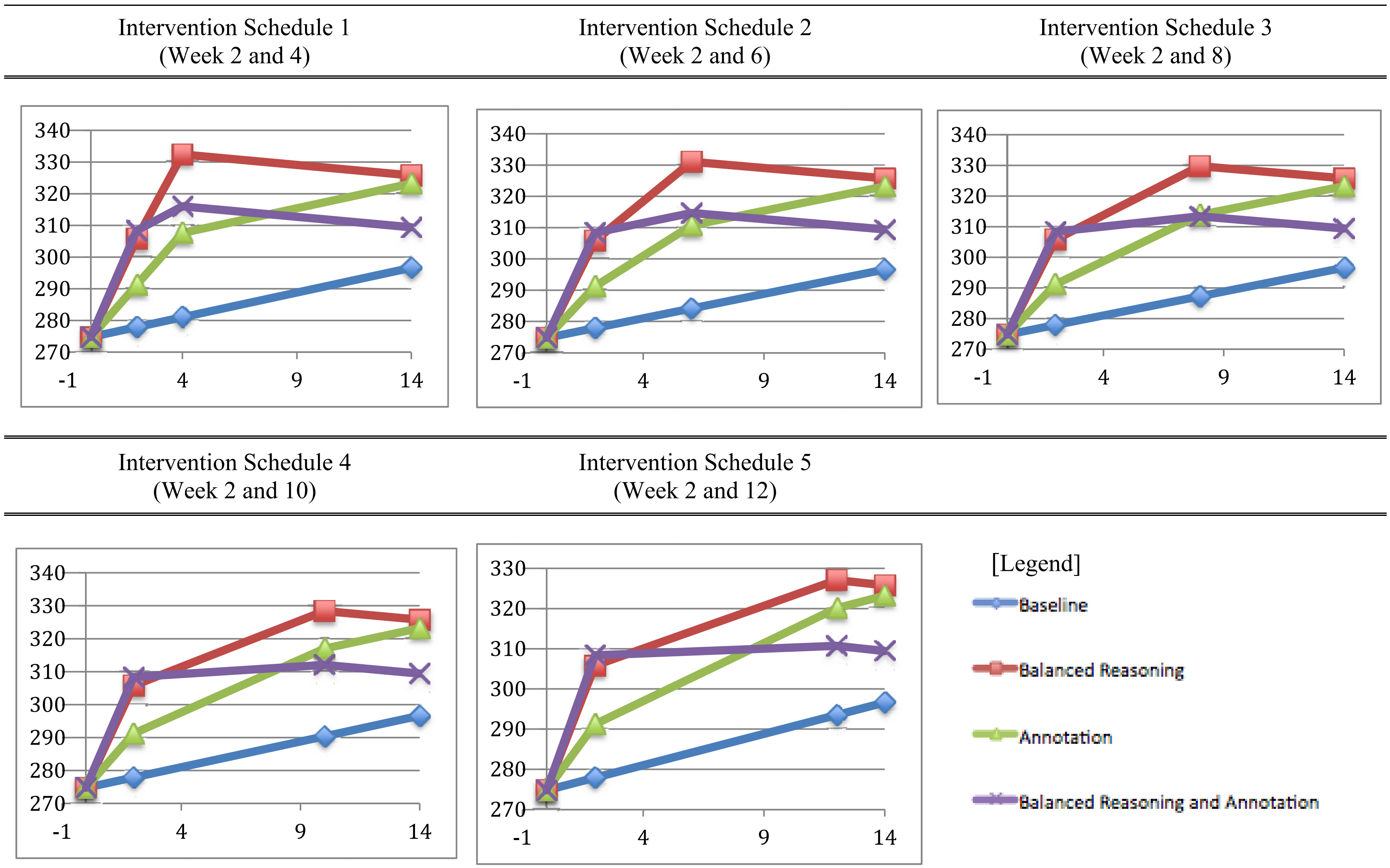

Figure 3.4. Predicted mean score change across intervention conditions and intervention schedules. The X and Y-axes respectively represent Time (weeks) and IEB scores. Each graph represents mean score change from differing intervention schedules.

Intervention Interaction. We expected synergistic, or enhancing effects of two treatment conditions combined (c.f., Tabak, 2004). For example, goal instructions would manipulate students' task perceptions during predictive evaluation and promote monitoring efforts afterwards. Annotation would structure source monitoring using advanced cues while reducing extraneous cognitive load. This assumption was partially supported. Each treatment condition added a positive differential to the normal growth slope. The elevation effect, however, was moderated by a negative interaction, which countervailed 26% of the possible sum of effects at the second measurement occasion. Students under combined treatment only slightly exceeded their peers. However, the interaction term functioned like a negative quadratic term; the score in the combined treatment condition subsequently stabilized.

These results might be attributed to a few causes documented previously. Balanced goals, asking students to consider conflicting views, might have reinforced instructional functions expected from the annotation tool (c.f., Wiley et al., 2009). Also, it is frequently assumed that simultaneous redundant scaffolds may void or even interfere by increasing extraneous cognitive load (Chandler & Sweller, 1996; Leslie, 2012; Mayer, Heiser, & Lonn, 2001). Though multiple, redundant scaffolding may benefit lower performance students (Brown et al., 1993; Tabak, 2004), they require well-integrated interfaces to negotiate between scaffolds. In our case, written instructions and separate tools use might have also interfered with each other. Students might have considered those interventions unimportant after repeated use.

Intervention-Time Interaction. We expected intervention effects to persist yet decay to a certain rate over time. Our findings partially supported this assumption. Source scaffolds and time interaction effects were statistically significant before considering level 2 predictors, but were non-significant in the final model. Perhaps source scaffolds had long-lasting effects on student evaluation by providing well-structured and repeated guidance over time. However, we noted a statistically significant, negative interaction of level 2 predictor WMC and time, which might have masked shared variation with source scaffolds.

Goal instructions and time interaction effects, on the other hand, were statistically significant but negative documenting the assumed decay of balanced goal effects over time. Despite higher IEB scores at 2nd and 3rd measurements, scores among students in the balanced goals condition ultimately stabilized. Thus, goal instructions proved simple and efficient, but their impacts were not fully sustained. Students might follow explicit instructions, but did not voluntarily activate critical perspectives in their final inquiries. These findings further support metacognitive researchers that address student perceptions about instructed strategies or interventions (Pressley et al., 1992): Learners need to know why specific strategies are important to sustain the impact.

The intervention-time interaction terms also helped to explain variability in growth patterns attributed to intervention schedules. The third task and corresponding interventions were scheduled at different times during the semester, ranging from 4-12 weeks. Students given their interventions within a two-week interval displayed sharper growth slopes between interventions, which also explained more gains at the second and third occasions among students given balanced goals or combined treatment before their reverse growth.

Implications, Limitations, and Closing Remarks

Implications for Practice

To appropriately support students, it is necessary to recognize that information evaluation requires complex metacognitive endeavors involving task interpretation, goal setting, multiple source representations with differing focuses, and thus working memory capacity for processing.

The current study supported the benefits of online annotation and balanced reasoning goals. Checklists provided simple and efficient support for lower performance students, but did not support advanced learners' needs (c.f., Harris, 2007; Meola, 2004; Metzger, 2007). As many researchers consistently suggest, scaffolding thus needs to be tailored to learners challenging and supporting learning (Pea, 2004; Puntambekar & Hübscher, 2005; van de Pol et al., 2010). Multiple scaffolds may well be needed to support intended performance. To ensure synergistic, enabling effects from combined scaffolds, however, instructors need to integrate scaffolds seamlessly while making those scaffolds visible and relevant to students. Students need to understand why and how given scaffolds may improve their performance (Clarebout & Elen, 2009; Hadwin & Winne, 2001; Sharma & Hannafin, 2004).

Methodological Implications

The study was conducted in naturalistic settings, which involved a quasi-experimental design. Instead of controlling for confounding variables, we employed multi-levels analysis to tease out variations attributed to pre-existing conditions and to model the specific nature of time-dependent changes. It is especially notable that we employed piecewise modeling strategies and time-varying intervention coding to better understand complex change patterns in the naturalistic settings. For instance, at the initial stage, students typically need time to be familiarized with new tools. Students also experience maturation through repeated use and stabilization after removal. We employed time-varying codes for interventions, and set separate intervention slopes and intervention-time interaction slopes. This modeling and coding approaches contain significant implications for the classroom intervention studies (c.f., Chou, Yang, Pentz, & Hser, 2004; McCoach & Kaniskan, 2010).

Directions for Future Research

The findings indicate that students' perceptions (e.g., perceived values and goals) and working memory capacity directly predict student's information evaluation behavior. From a metacognitive perspective, requisite processing is readily assumed, but has rarely been critically tested for online information seeking and evaluation (c.f., Lazonder & Rouet, 2008; Rieh & Hilligoss, 2008). Considering prevalent online inquiries and evaluation tasks in everyday and academic tasks, further research should enhance our understanding of dynamics among students, information, and tasks (c.f., Kuiper et al., 2005). For instance, existing task perceptions likely influence when students consider potentially biased information available on the Web and social media.

Also, the results indicate that statistically significant variability was not explained in our final model. Further examination of individual variables is needed to design, refine, and sustain the effects of scaffolds. For example, researchers suggest the tendency toward deep thinking (i.e., need for cognition) influence information-seeking (e.g., Mason, Boldrin, & Ariasi, 2009). If so, these conclusions may well be linked to information evaluation behavior.

We also acknowledge certain limitations. We examined evidence from four repeated measures, which limited closer scrutiny of interaction between scaffolds and individual differences. Thus, the scaffold effects might be over- or under-estimated. For example, perceived task value and working memory might interact with scaffold use differentiating scaffolds effects. The online annotation effect might have varied over time. Further examinations of these interaction effects are needed. Also, we relied on self-reports to examine change patterns and to test the impact of scaffolds. By employing advanced analytical strategies, we examined various aspects of student information evaluation behaviors, but repeated studies are needed to triangulate data source and address limitations using complementary inquiry methods.

In closing, recent surveys document drastic changes in task types to inquiry or research and subsequent challenges students confront during their transition to college (Alison & Michael, 2009). Information evaluation is increasingly important part of academic inquiries and involves complex mental tasks, which may account for certain academic challenges college students face. We identified again the need and examined varying solutions, for additional support for college students especially differentiating information evaluation skills and activated behaviors.

REFERENCES

Apedoe, X. S., & Reeves, T. C. (2006). Inquiry-based learning and digital libraries in undergraduate science education. *Journal of Science Education and Technology*, *15*(5), 321–330.

Asher, A. D. (2011). Search magic: Discovering how undergraduates find information. Presented at the American Anthropological Association Annual Meeting, Montreal, Canada.

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2010). Problem-based learning and argumentation: testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*. doi:10.1007/s11251-010-9148-z

Belland, B. R., Glazewski, K., & Richardson, J. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research & Development*, *56*(4), 401–422. doi:10.1007/s11423-007-9074-1

Borkowski, J. G., Carr, M., & Pressley, M. (1987). "Spontaneous" strategy use: Perspectives from metacognitive theory. *Intelligence*, *11*(1), 61–75. doi:16/0160-2896(87)90027-4

Bowler, L. (2010). A taxonomy of adolescent metacognitive knowledge during the information search process. *Library & Information Science Research*, *32*(1), 27–42. doi:16/j.lisr.2009.09.005

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning & Instruction*, *21*(2), 180–192. doi:10.1016/j.learninstruc.2010.02.002

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*(2/3), 191–213.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition & Instruction*, *20*(4), 485.

Brown, A., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188–228). New York, NY: Cambridge University Press.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13–27. doi:10.1016/j.learninstruc.2007.12.003

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151–170.

Cho, K.-L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, *50*(3), 5–22.

Chou, C.-P., Yang, D., Pentz, M. A., & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis*, *46*(2), 213–225. doi:10.1016/S0167-9473(03)00149-X

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ (Clinical Research Ed.)*, *324*(7337), 573–577.

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, *92*(4), 694–702. doi:10.1037/0022-0663.92.4.694

Gama, C. (2004). Metacognition in interactive learning environments: The reflection assistant model. Springer.

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, *46*(2), 127–146.

Hannafin, M. J., Hannafin, K., & Gabbitas, B. (2009). Re-examining cognition during student-centered, web-based learning. *Educational Technology Research and Development*, *57*(6), 767–785. doi:10.1007/s11423-009-9117-x

Harris, F. J. (2007). Challenges to teaching credibility assessment in contemporary schooling. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 155–180). Cambridge, MA: The MIT Press. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/dmal.9780262562324.155

Head, A., & Eisenberg, M. (2009a). *Finding context: What today's college students say about conducting research in the digital age*. University of Washington's Information School. Retrieved from http://projectinfolit.org/pdfs/PIL_ProgressReport_2_2009.pdf

Head, A., & Eisenberg, M. (2009b). Lessons learned: How college students seek information in the digital age. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2009_finalv_YR1_12_2009v2.pdf

Head, A., & Eisenberg, M. (2010). Truth be told: How college students evaluate and use information in the digital age. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2010_Survey_FullReport1.pdf

Hjørland, B. (2012). Methods for evaluating information sources: An annotated catalogue. *Journal of Information Science*, *38*(3), 258–268. doi:10.1177/0165551512439178

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist*, *39*(1), 43–55.

Hogarth, R. M. (1980). *Judgment and choice: The psychology of decision*. John Wiley & Sons Inc.

Hughes, B., Wareham, J., & Joshi, I. (2010). Doctors' online information needs, cognitive search strategies, and judgments of information quality and cognitive authority: How predictive judgments introduce bias into cognitive search models. *Journal of the American Society for Information Science & Technology*, *61*(3), 433–452. doi:10.1002/asi.21245

Iding, M. K., Crosby, M. E., Auernheimer, B., & Barbara, E. (2008). Web site credibility: Why do people believe what they believe? *Instructional Science*, *37*, 43–63. doi:10.1007/s11251-008-9080-7

Jiang, L., & Elen, J. (2011). Why do learning goals (not) work: a reexamination of the hypothesized effectiveness of learning goals based on students' behaviour and cognitive processes. *Educational Technology Research and Development*, *59*(4), 553–573. doi:10.1007/s11423-011-9200-y

Jones, B. D. (2001). Recommendations for implementing Internet inquiry projects. *Journal of Educational Technology Systems*, *30*(3), 271–291.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. doi:10.1007/s10648-007-9054-3

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410. doi:16/j.learninstruc.2009.02.003

Kalyuga, S. (2011). Informing: A cognitive load perspective. *Informing Science*, *14*, 33–45.

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*(3), 209–215. doi:10.1007/s11251-009-9102-0

Kirk, R. E. (1994). *Experimental design : procedures for the behavioral science* (3rd ed.). Wadsworth Publishing.

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services* (2nd ed.). Westport, CT: Libraries Unlimited.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind.* (pp. 301–326). Hove England: Psychology Press.

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, *75*(3), 285–328.

Lan, W. (2005). Self-monitoring and its relationship with educational level and task importance. *Educational Psychology*, *25*(1), 109–127. doi:10.1080/0144341042000294921

Lazonder, A. W., & Rouet, J.-F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, *24*(3), 753–765. doi:10.1016/j.chb.2007.01.025

Lee, Y.-J. (2004). Concept mapping your Web searches: a design rationale and Web-enabled application. *Journal of Computer Assisted Learning*, *20*(2), 103–113. doi:10.1111/j.1365-2729.2004.00070.x

Leslie, K., RenaeJin, PutaiSweller, John. (2012). Redundancy and expertise reversal effects when using educational technology to learn primary school science. *Educational Technology Research & Development*, *60*(1), 1–13. doi:10.1007/s11423-011-9199-0

Mason, L., Boldrin, A., & Ariasi, N. (2009). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607–633. doi:10.1007/s11251-008-9089-y

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, *5*(1), 67–90.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. doi:10.1037/0022-0663.93.1.187

McCoach, D. B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in Quantitative Psychology and Measurement*, *1*, 17. doi:10.3389/fpsyg.2010.00017

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy*, *4*(3), 331–344.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. doi:10.1002/asi.20672

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education*, *41*(3), 271–290. doi:10.1016/S0360-1315(03)00049-6

Miyata, Y., & Norman, D. A. (1986). The control of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design : new perspectives on human-computer interaction* (pp. 265–284). Hillsdale, N.J.: L. Erlbaum Associates.

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

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. doi:10.1037/0022-0663.97.2.157

Oh, S., & Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning*, *23*(2), 95–110. doi:10.1111/j.1365-2729.2006.00206.x

Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, *91*(2), 230–240. doi:10.1037/0022-0663.91.2.230

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What Research Has to Say About Reading Instruction* (pp. 291–309). Newark, DE: International Reading Association.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. *Annals of Child Development*, *4*, 89–129.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, *92*(5), 513–555.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science & Technology*, *53*(2), 145–161.

Rieh, S. Y., & Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 49–71). Cambridge, Mass: MIT Press.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition & Learning*, *2*(2/3), 125–140. doi:10.1007/si1409-007-9010-0

Rouet, J. (2006). *The skills of document use: From text comprehension to Web-based learning*. L. Erlbaum Associates.

Rouet, J.-F. (2009). Managing cognitive load during document-based learning. *Learning and Instruction*, *19*(5), 445–450. doi:16/j.learninstruc.2009.02.007

Russell, D. M., Stefik, M. J., Pirolli, P. L., & Card, S. K. (1993). *The cost structure of sensemaking* (pp. 269–276). Presented at the ACM SIGCHI conference on human factors in computing systems.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science & Technology*, *58*(13), 2126–2144. doi:10.1002/asi.20681

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition□. *Information Processing & Management*, *26*(6), 755–776. doi:10.1016/0306-4573(90)90050-C

Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science*, *38*(5), 487–502.

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah, N.J: Lawrence Erlbaum Associates.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*(1-2), 113–25.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–75.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, *2*(2-3), 191–210.

Stadtler, M., Scharrer, L., & Bromme, R. (2011). How reading goals and rhetorical signals influence recipients' recognition of intertexual conflicts. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Sceicne Society* (pp. 1346–1351). Austin, TX: Cognitive Sciecne Society.

Stromoso, H. I., & Braten, I. (2002). Norwegian law students' use of multiple sources while reading expository texts. *Reading Research Quarterly*, *37*(2), 208–227.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, *13*(3), 305–335.

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research (07408188)*, *27*(2), 222–248. doi:10.1016/j.lisr.2005.01.005

Tanni, M., & Sormunen, E. (2008). A critical review of research on information behavior in assigned learning tasks. *Journal of Documentation*, *64*(6), 893–914.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. doi:10.1016/0749-596X(89)90040-5

Walraven, A., Brand-gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, *24*(3), 623–648. doi:doi: 10.1016/j.chb.2007.01.030

Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. doi:10.1006/ceps.1999.1015

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346. doi:10.1016/S0364-0213(99)80043-3

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York: Routledge.

Wu, Y.-T., & Tsai, C.-C. (2005). Information commitments: Evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, *21*(5), 374–385. doi:10.1111/j.1365-2729.2005.00144.x

Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, *58*(1), 181–196.

CHAPTER 4

THE EFFECTS OF SOURCE REPRESENTATIONS AND GOAL INSTRUCTIONS ON COLLEGE STUDENTS' EVIDENCE QUALITY, REASOSNING, AND CONCEPTUAL INTEGRATION IN ARGUMENTATION-BASED INQUIRY[4]

[4] Kim, S. M., & Hannafin, M. J. To be Submitted to *Instructional Science*

Abstract

This study examined how differing instructional interventions—source representation scaffolds and goal instructions— influence the actual use of evaluation skills to improve argumentation quality. Group and individual writing were examined during and after interventions respectively. Students' scaffolds use experience and cognitive load were examined. The results were mixed. Goal instructions did not have direct effects on argumentation quality while the source representation scaffolds had significant effects only when given balanced reasoning goals or mixed-prior knowledge group dynamics. The results supported the synergistic integration of two scaffolding functionality, yet suggested addressing possible difficulties in using scaffolds and sustaining the effects in the complex classroom situations.

Keywords: information evaluation, goal instructions, source representation scaffolds

Introduction

Online inquiry and argumentative writing constitute important academic tasks. Argumentation involves building valid answers to questions when given incomplete or incompatible information (Driver, Newton, & Osborne, 2000; Hirsch, 1989; Jonassen & Kim, 2010). Effective argumentation relies on evidence quality, grounded reasoning, and conceptual integration (Brem & Rips, 2000; Clark & Sampson, 2008). College entering students, however, often report challenges in conducting research to define, search, integrate, and present needed information (see for a review, Head, 2013), and tend to engage in pseudo-inquiry involving *fact-finding* (c.f., Marchionini, 2006). This tendency is characterized by reliance on safe, instructor-proved information sources (Head & Eisenberg, 2009), confirmation bias (Nussbaum & Kardash, 2005), unsubstantiated or unwarranted explanation (Brem & Rips, 2000; Kuhn, 1989), and too-little or too-much information without elaboration (Asher, 2011).

On the other hand, information evaluation is a critical component of successful inquiry and argumentation. Information evaluation involves a series of metacognitive tasks to *reflect and control one's knowledge sources*. For example, one may consider whether a document is credible and compatible with competing alternatives to support a given claim. Such higher-order, source property representations and their inter-links (e.g., "this peer-reviewed source disagrees with prevalent views") are indexed to emerging mental models to influence associated courses of action such as filtering, re-reading, marking and linking source distinctions, and seeking additional sources (c.f., Pressley, 2002; Rouet, 2006; Schraw, 1998; Winne & Nesbit, 2009). Effective evaluation thus orients unfamiliar information seeking (Tabatabai & Shore, 2005; Walraven, Brand-gruwel, & Boshuizen, 2008), helps to detect rhetorical linkages and qualifiers across multiple documents (Perfetti, Rouet, & Britt, 1999), and contributes to richer mental model construction (Bråten, Strømsø, & Britt, 2009; Wiley et al., 2009), therefore possibly contributing to quality argumentation (c.f., Clark & Sampson, 2008).

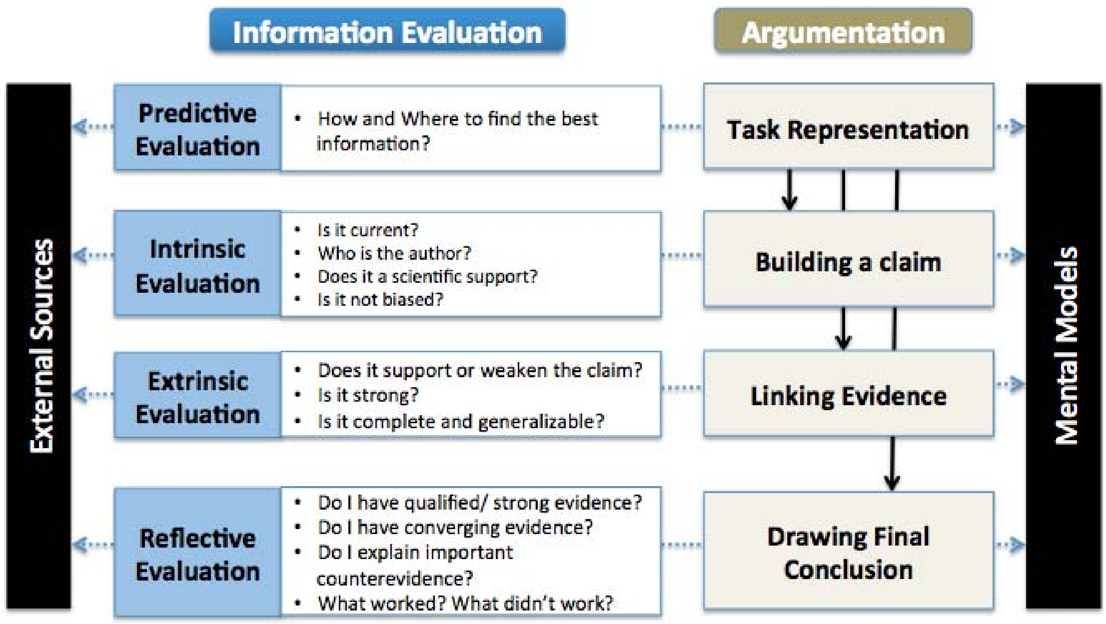

Figure 4.1. Information evaluation during argumentation model. Adapted from Kim & Hannafin (2013)

Figure 4.1 depicts how continual information evaluation is interlinked with, and crucial to, argumentation while representing tasks, building a claim, linking evidence to the claim, and drawing a conclusion. Information evaluation initially responds to task specifics and evolves through predictive, intrinsic, extrinsic, and reflective evaluation. First, predictive evaluation (e.g., "Wikipedia is efficient but not valid to build a substantiated knowledge base") sets expectations, or standards about needed information sources, to influence information-seeking scope and strategies (c.f., Hughes, Wareham, & Joshi, 2010; Wu & Tsai, 2005). Intrinsic evaluation focuses on the innate merit of sources (e.g., currency, support, authority, and bias) (Rieh & Hilligoss, 2008; Wineburg, 1998). Intrinsic evaluation contributes to locating key information sources while reducing extraneous cognitive load associated with initial trials and errors (c.f., Kalyuga, 2011). Extrinsic evaluation involves weighing the strength of competing sources to support claims (Hjørland, 2012; Meola, 2004). Extrinsic evaluation helps to represent linkages and qualifiers across multiple sources (Perfetti et al., 1999), thus enables logical source integration and coherent explanations (Bråten et al., 2009; Wiley et al., 2009). Finally, reflective evaluation involves summative reflection across collected sources to strengthen global linkages among sources while reconciling and reorganizing unexplained, often contradictory sources (c.f., Russell, Stefik, Pirolli, & Card, 1993).

Seasoned information evaluation, however, is rarely activated due to simplified task perceptions and low evaluation standards (Wu & Tsai, 2005), mechanical and superficial level cue recognition (Bråten et al., 2009; Britt & Aglinskas, 2002), and high cognitive load from multiple information seeking tasks (Eysenbach & Köhler, 2002; Kalyuga, 2009; Rouet, 2009). In a previous study with introductory college biology students (Kim & Hannafin, 2014), we integrated scaffolds: a) to help represent complex source properties across sources while compensating for working memory capacity (source representation scaffolds); b) to manipulate goals to induce higher information evaluation standards and positive predictive evaluation (goal instructions). Source representation scaffolds included a checklist (control) and an online annotation tool designed to externalize and integrate source representation within the argumentation process (treatment). Goal instructions included a general persuasion goal ("persuade or convince the public of your claim") in the control condition and a balanced reasoning goal ("critically compare various claims" or "provide the reason why your counterarguments are wrong") in the treatment condition. Self-reports indicated treatment conditions helped students to increase attention to varying source properties and improve information evaluation behavior over time. In the current study, we further examined whether and how treatment conditions influenced student information use for argumentative writing: evidence quality, reasoning, and conceptual integration in student writing. We addressed three research questions:

1. What are the intervention effects of source representation scaffolds and goal instructions on group argumentation quality?
2. What are the transfer effects of source representation scaffolds and goal instructions on individual argumentation quality?
3. How do students respond to goal instructions and source representation scaffolds during their argumentation-based inquiry?

Research Method

We implemented mixed-methods research (Teddlie & Tasshakori, 2006) to converge quantitative and qualitative data to provide evidence of the influence of scaffolding on argumentation. Table 4.1 aligns research questions with design, data collection methods, and analytic strategies. For research questions one and two, quasi-experimental studies examined scaffolds effects using MANCOVAs. We expected treatment conditions would benefit students in argumentation; we also expected there would be positive interaction effects meaning synergistic relations between goals and source representation. Research question three examined how students used the scaffolds to determine the influence of the interventions. Responses to survey items were analyzed both quantitatively and qualitatively to compare task perceptions, source representation tools use, and cognitive load across conditions. We generated final inferences based on the results from each strand. Figure 4.2 summarizes the current study. The specific methods, techniques and results for each research question are detailed in the following sections.

Table 4.1

Alignment of Research Questions, Design, Data Collection, and Analytic Strategies

Research Question	Research Design	Data Collection	Data Analysis
1. What are the intervention effects of source representation scaffolds and goal instructions on group argumentation quality?	• Quasi-experimental study (N=99)	• Group argumentation quality (Group writing rubric scores—evidence quality, reasoning, and conceptual integration)	• MANCOVA
2. What are the transfer effects of source representation scaffolds and goal instructions on individual argumentation quality?	• Quasi-experimental study (N=482)	• Individual argumentation quality (Individual writing rubric scores—evidence quality, reasoning, and conceptual integration)	• MANCOVA
3. How do students respond to goal instructions and source representation scaffolds during their argumentation-based inquiry?	• Survey (N=482)	• Task perceptions (open response item)	• Thematic analysis • Cluster analysis
	• Survey (N=482)	• Source representation tools use (open response item)	• Thematic analysis
	• Survey (N=464)	• Source representation tools cognitive load (NASA-Task Load Index)	• T-test

Note. Final sample size was 482; however, due to non-response, cognitive load was measured and examined for 464 students (96%).

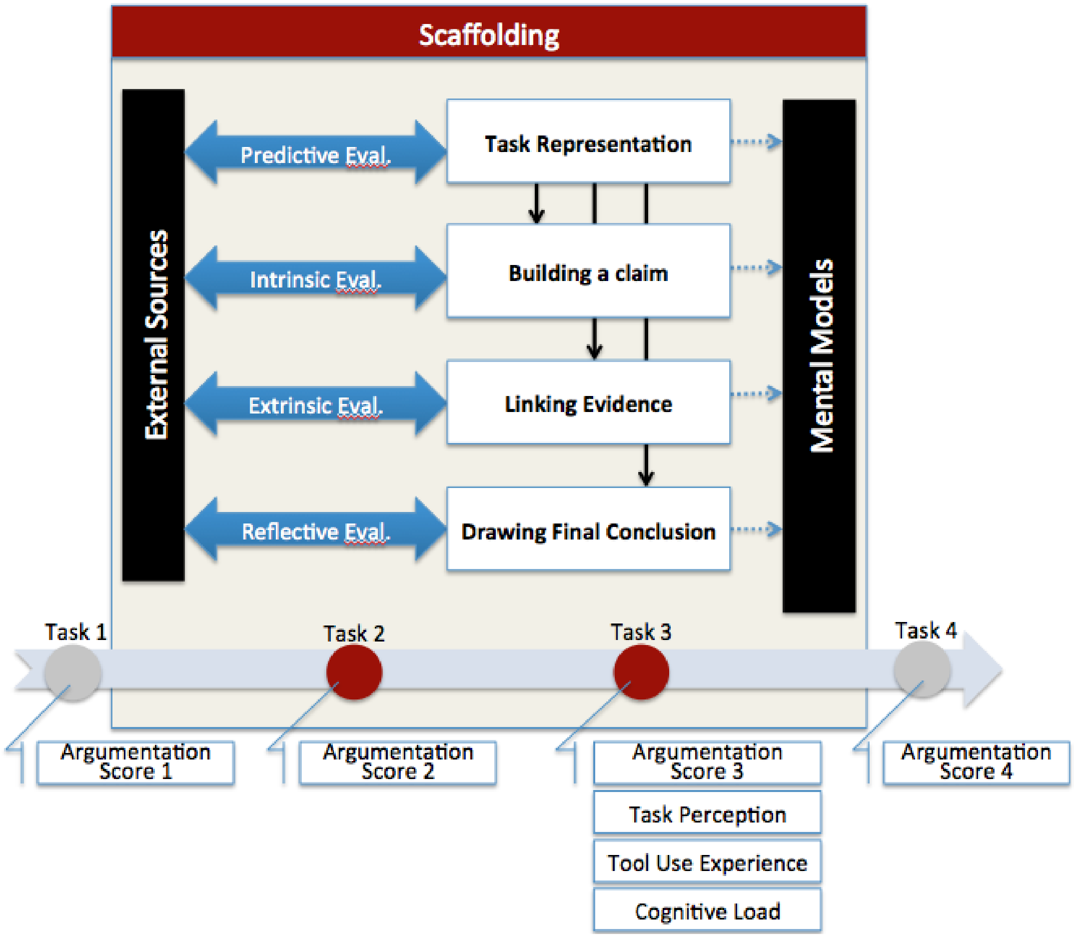

Figure 4.2. Study overview

Settings and Participants

Two intact sections from the same college biology course participated in this study in 2012 Spring. The course was a general education requirement for non-major students at a large public university in the southeast United States. The same instructor incorporated online inquiry throughout the semester. The instructor expressed interest in devising effective and efficient approaches to scaffold students' information evaluation and use during projects,

The study included 488 students who completed all the inquiry task requirements for individual analysis; 6 multivariate outliers were eliminated based on significant Mahalanobis Distance (P<.001). The final sample therefore included 482 students: 73% female, 74%

freshmen, and 87% non-science majors according to a demographics survey (see Appendix A). On a prior knowledge test of important course concepts, 91% students scored 50% or lower (mean score=6.31; maximum possible=18) (see Appendix B for the test items). This ensured that students did not already demonstrate the target knowledge applied during inquiry tasks. Overall, we included 99 small groups who worked throughout the semester without changing group membership. Group size varied from 2 to 8 with mean size 6. Students formed self-selected groups with differing levels of prior knowledge: 33% groups were low-prior knowledge groups (all members below 50%); the remaining 67% were mixed-prior knowledge groups.

Materials and Procedures

All students were exposed to the checklist use as part of classroom interactions. Each section was then assigned to either checklist only (control) or online annotation tool + checklist condition (treatment) (factor A). Project groups were randomly assigned either to a persuasion goal condition (control) or balanced reasoning goal condition (treatment) (factor B). Table 4.2 summarizes the number of students assigned to each condition.

Table 4.2

Interventions Assignment Scheme

		Factor B	
		Persuasion Goals(C)	Balanced Reasoning Goals(T)
Factor A	Checklist (C)	24 (125)	25 (104)
	Checklist + Annotation (T)	25 (123)	25 (130)

Note. Cell numbers represent groups; parenthetical numbers represent individual students. C and T represent control and treatment conditions respectively.

Inquiry Tasks. The course emphasized an inquiry curriculum, where students used the Web resources to examine different aspects of biological phenomena to dispute prevalent misconceptions. For example, students were asked to write a newspaper health column that explains which molecules build up in muscle cells causing them to burn during exercise. On another project, students chose a vaccine and examined different claims in regard to its safety. Students conducted one common group task and then the other chosen group task among six topics (refer to Appendix C for the list of topics). Before and after group tasks, students engaged in two individual tasks, which measured individual students' entry and final performance.

During week 2, students conducted the first individual task without interventions. Students were then trained to use the checklist or/and the online annotation tool. Interventions were provided along with the second and third group tasks and removed during the final individual task. Immediately after each task, students responded to post-task surveys asking their perceptions and experiences. Specific interventions are detailed below.

Source representation scaffolds. The written checklist (control), the most commonly used aid to support information evaluation, included evaluation criteria students could use to examine information sources. The criteria included CRAP—currency, reliability, author, and bias in a low contextual manner (Britt & Aglinskas, 2002; Wiley et al., 2009). The instructor continued to emphasize CRAP testing during class sessions to automatize information evaluation. Students, individually and in groups, submitted source evaluation rationale after each group task.

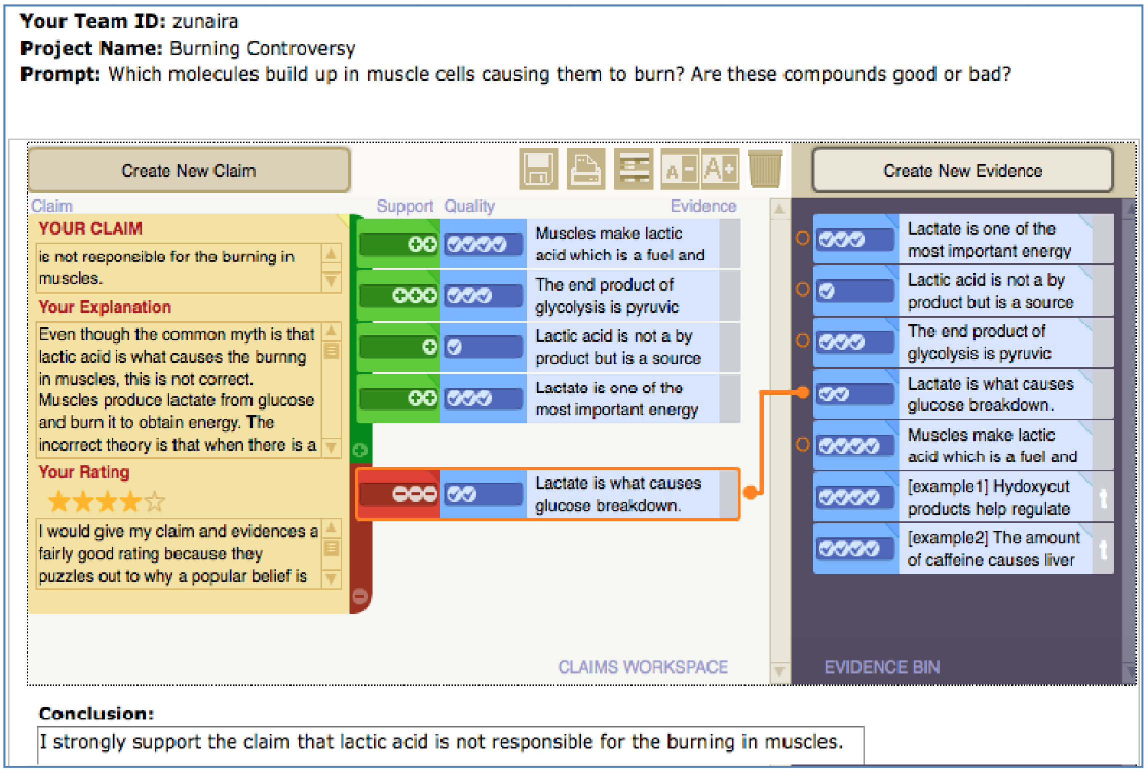

Figure 4.3. Annotation tool screen shot. Students completed six steps: (a) Creating claim; (b) creating evidence; (c) rating evidence quality (intrinsic evaluation); (d) rating evidence strength (extrinsic evaluation); (e) rating the strength of claim (reflective evaluation); and (f) drawing a final conclusion.

In contrast, the online annotation tool—*Showing Evidence*[5]—provided a visual framework for identifying, judging, and linking evidence that supports or weakens claims within argumentation structure (see Figure 4.3). Students in the treatment used the annotation tool to pool documents from the Internet search and integrate them to their argumentation. They either established their own claim from the beginning or created a temporary claim by scanning collected documents. Students registered each document in the evidence bin to record and judge source properties (e.g., author, date, and abstract) using a 5-point rating system for intrinsic quality (e.g., It is basically about OOO and is from a qualified source because..."). Next, they visually linked each source to their claim and judged how strongly identified evidence supported

5 The authors acknowledge Intel® Innovation in Education for permission to use the tool in this research.

or weakened their claim using a 5-point rating system and documented their reasoning (e.g., "I believe this strongly weakens my claim since…") as a warrant statement for their argumentation. Students finally evaluated all information collected (e,g., "is my collection integrated into a consistent whole?") to rate the strength of their claim and to draw a final conclusion. Individual annotations were shared to create a group annotation, thereby creating a final group paper.

Though simple and efficient, checklists focused on intrinsic evaluation, required evaluation as a separate task from source integration into arguments, and rarely addressed cognitive hurdles (e.g., task perceptions or cognitive load) student might face. Thus, it was expected to limit holistic representation across multiple sources. On the other hand, the annotation tool assisted in documenting information externally, thereby allocating working memory effectively during information seeking and evaluation (c.f., Belland, Glazewski, & Richardson, 2008; Lee, 2004; Miyata & Norman, 1986; Zhang & Quintana, 2012). It modeled and integrated advanced cues for extrinsic and reflective evaluation within an argumentation task in a high contextual manner (c.f., Cho & Jonassen, 2002; Oh & Jonassen, 2007; Reiser, 2004).. Therefore we expected that students who used the annotation tool would demonstrate more quality evidence, logical reasoning, and conceptual integration in their writing. However, we expected that the cognitive load students perceived would be slightly higher than the checklist since the tool involved intrinsically more complex evaluation tasks that the checklist did not necessitate.

Goal instructions. To induce higher information evaluation standards and positive predictive evaluation, written goal statements were embedded in the second and third group task instructions. The persuasion goal (control) asked students to write a persuasive letter or convince their peers to *accept their claim*. The balanced reasoning goal (treatment) asked students to *support an informed decision* of the audience while critically examining opposing arguments and multiple points of view. For the first and fourth individual tasks, we used a neutral statement, "write your essay based on your understanding" (refer to Appendix F for example statements).

We expected simply changing the specificity of written directions for a given task can influence students' task perceptions and behavior (Ferretti, MacArthur, & Dowdy, 2000; Nussbaum & Kardash, 2005; Page-Voth & Graham, 1999). However, literature cautioned that a general persuasion goal ("convince or persuade your friend about your claim"), which is the most prevalent form of science task directions, might induce students' efforts to ignoring counter-arguments. On the other hand, we expected that balanced reasoning goals ("is there any significant counter-evidence left unexplored?") to help students to define and seek alternative sources appropriate to their task (Jiang & Elen, 2011; Winne & Hadwin, 1998). Therefore, balanced goals would direct attention to those sources (Song, 2005; Locke & Latham, 2002), induce efforts to seek global coherence across competing sources(Stadtler, Scharrer, & Bromme, 2011), and thus contribute to better understanding (Wiley et al., 2009).

Argumentation Quality

Following suggestions for argumentation (see for a review, Clark & Sampson, 2008; Kuhn, 1989) and learning outcome assessment (Biggs & Collis, 1982; Chi, Slotta, & De Leeuw, 1994), we developed three rating rubrics to measure argumentation quality (see Appendix L). First, e*vidence quality* measured the quality and strength of sources in the in-text citations and reference list in a summative manner (scores 0 to 5): Level 0 (no evidence), level 1 (unqualified evidence), level 2 (qualified evidence), level 3 (converging evidence), level 4 (alternative evidence), and level 5 (dialectic evidence). The final level meant students considered both agreeing and opposing evidence with a rebuttal to create a strong argument. R*easoning* measured rhetorical signals in writing to see how explicitly students use their source representations to incorporate differing sources of evidence and create grounded arguments (scores 0 to 4): Level 0 (unjustified), level 1(unsubstantiated), level 2(substantiated), level 3-4 (warranted). Students assigned to level 4 substantiated their claims explicitly citing, qualifying, and relating multiple evidences. Finally, *conceptual integration* measured conceptual elaboration in writing. Since two individual task topics incorporated similar conceptual structures—kinds of substances (e.g., plastics), cause (e.g., heat), and process (e.g., leaching and estrogen mimics), we initially developed and applied an analytic rubric summing content specific scores (summed scores 0 to 5). However, due to variability across seven differing group task topics, we adopted SOLO taxonomy (Structure of Observed Learning Outcomes) (Biggs & Collis, 1982) to classify group writing (scores 0 to 4) : Level 0 (pre-structural), level 1 (uni-structural), level 2 (multi-structural), level 3 (relational), and level 4 (abstract).. Advanced levels meant students connected multiple points in a correct context (level 3); students theorized or evaluated alternative arguments properly (level 4).

Data Collection and Analysis

Group and Individual Argumentation Quality. In order to examine scaffolding effects on argumentation quality, group and individual writing was analyzed. Two raters—doctoral students in science domains blind to intervention conditions—applied three rubric schemes we created. After scoring training with 30 writing sample, they individually rated common writing samples—randomly selected 10% individual writing (N=100) and 100% group writing (N=198) samples. The first author examined major differences in scores, and raters continued re-rating till they reached close agreement (Cohen's Kappa statistics=.95 on average). The remaining writing samples were split and measured independently.

Data were analyzed using Multivariate Analysis of Covariance (MANCOVA). We used the initial individual or group task scores as covariates, the final individual or group task scores as performance variables, and source representation and goal instructions as main factors. Since we expected group composition (low vs. mixed-prior knowledge groups) would moderate group performance and the subsequent individual performance (Belland, 2010; Carter & Jones, 1994; White & Frederiksen, 1998), group composition was entered as a moderator. Main tests were carried out at an alpha level .05, which was adjusted for follow-up tests using the Bonferroni method.

Scaffolds Experience. We conducted surveys to examine how and why students responded to varying scaffolding conditions. More specifically, we examined task perceptions in response to goal instructions, source representation tools use experience, and cognitive load.

Task perceptions. As part of post-task surveys, we asked students to freely describe their task perceptions: "You've read this task direction ___. What did you think this task was all about? What was your concern? How did you expect to deal with it?" Nineteen categories (e.g., "Fixed-Science Goal) of task attributes were identified through an iterative thematic analysis (see Appendix M for the list). The first author and another doctoral student applied binary codes to student responses with 1 indicating each category occurrence to create individual perception profiles (an average Cohen's Kappa=.81). For example, one student was coded as having an *open-science goal (*expecting controversy and arguments) with a clear *evaluation focus* and *extrinsic evaluation* strategies*.* Another student was coded as having a *fixed-science goal* with a focus on content *learning*. We then conducted a Two-Step Cluster Analysis to summarize individual profiles to reveal natural groupings in task perceptions when given persuasion or balanced goal instructions.

Source representation tools use. Another open-ended question as part of the final survey asked students to describe their source representation tool use experience: "How did you use the annotation tool (or checklist) during your inquiry? Which aspects helped or restricted your task completion?" We conducted thematic analysis to examine possible differences between the checklist and annotation tool use.

Source representation tool cognitive load. NASA-Task Load Index (NASA-TLX) was administered to examine perceived cognitive load during annotation or checklists use. NASA-TLX has been reported highly correlated with other cognitive load measures (Hart & Staveland, 1988) while increasing interpretability with its multidimensional scale. The original scale consists of six items, among which we used four items appropriate for the current study context—mental demand, time pressure, efforts, and frustration to create a 100-point scale. The reliability was .72. We summed up sub-scales to create an index for a t-test. Possible scores ranged from 0 to 400.

Results and Discussion

Effects of Source Representation Scaffolds and Goal Instructions

We examined final writing artifacts (N=482) from each group to examine effects of source representation scaffolds and goal instructions on argumentation while controlling for previous group argumentation scores. As expected, the initial three-way MANCOVA indicated significant two-way and three-way interaction effects on overall group argumentation. A source-goal interaction effect, Λ=0.90, $F(3,86)=3.36$, $p= .023$, $\eta_p^2= .11$, indicated that the source representation main effect varied depending on goal instructions and vice versa; this interaction effect again varied depending on group compositions (Λ=0.88, $F(3,86)=3.97$, $p=.011$, $\eta_p^2=.12$). Therefore, we decomposed three-way interaction to test simple interaction effects at different group-composition conditions(Howell & Lacroix, 2012); alpha level was adjusted to test two contrasts (.05/2=.025). The source-goal interaction was only significant among low-prior knowledge groups (Λ= 0.87, F(3,86)=4.43, p=.006, η_p^2=.13). Given different dynamics among factors, we split data to conduct separate analyses between low-prior knowledge (N=33) and mixed-prior knowledge groups (N=66). Table 4.3 and Figure 4.4 contrast results from the analyses across group composition conditions.

Low-Prior Knowledge Groups. Given the significant source-goal interaction, we tested simple main effects of scaffolding (an alpha level =.025/2=.0125). Results indicated the source representation effect was statistically significant only under balanced goal instructions (Λ= 0.51, F(3,24)=7.61, p=.001). The effect size indicated that source representation scaffolds explained 49% variability of overall group argumentation quality. We then conducted three follow-up ANOVAs (alpha level =.0125/3=.004) to test simple effects on each univariate dependent variable. As Table 3 illustrates, statistically significant differences were detected for evidence quality (F(1,26)=11.49, p=.002, η_p^2=.31, *d*=1.57) and reasoning (F(1,26)=7.30, p<.001, η_p^2=.44, *d*=2.40). The adjusted mean score of evidence quality (4.14) indicated that students incorporated counter-evidence. Reasoning (2.93) indicated that student writing approached the performance of warranted arguments; student writing included qualifiers ("a *current peer-reviewed* article *strongly* suggested…") or linkages among sources ("in support of this claim, *converging* evidence indicated …across fields").

Table 4.3

Source Representation Simple Effects on Group Argumentation across Group Composition Conditions

Group	Dependent Variable	Condition	N	Group Task1		Group Task2		MANOVA				ANOVA			
				M	SD	Adj.M	SD	Λ	F	p	η_p^2	F	p	η_p^2	*d*
Low-prior Knowledge Groups (N=33)	Evidence Quality	Checklist	15	3.29	1.38	2.67	.82	.51	F(3,24) =7.61	.001	.49	F(1,26) =11.49	.002	.31	1.57
		Annotation	18	3.71	.95	4.14	.93								
	Reasoning	Checklist	15	3.00	1.00	1.42	.63					F(1,26) =20.53	.000	.44	2.40
		Annotation	18	2.57	.54	2.93	.63								
	Conceptual Integration	Checklist	15	1.43	.79	2.40	.85					F(1,26) =.03	.867	.001	.08
		Annotation	18	2.00	1.00	2.47	.82								
Mixed-prior Knowledge Groups (N=66)	Evidence Quality	Checklist	32	3.09	.90	2.67	.93	.70	F(3,57) =8.06	.000	.30	F(1,59) =17.79	.000	.23	1.04
		Annotation	34	2.88	.98	3.63	.91								
	Reasoning	Checklist	32	2.62	.78	2.37	.76					F(1,59) =14.10	.000	.19	.91
		Annotation	34	2.22	.82	3.05	.73								
	Conceptual Integration	Checklist	32	1.68	1.04	2.44	.99					F(1,59) =2.82	.098	.05	.41
		Annotation	34	1.75	.92	2.85	1.02								

Note. Low-prior knowledge group analysis indicates simple effects of source representation given balanced reasoning goal instructions while mixed-prior knowledge group analysis indicates source representation effects independent of goal instructions conditions. Evidence quality 0 (no evidence), 1 (unqualified evidence), 2 (qualified evidence), 3 (converging evidence), 4 (alternative evidence), 5(dialectic evidence); Reasoning 0 (unjustified), 1(unsubstantiated), 2(substantiated), 3-4 (warranted); Conceptual integration 0 (pre-structural), 1 (uni-structural), 2 (multi-structural), 3 (relational), 4 (abstract)

(a) Low-Prior Knowledge Groups (Given Balanced Reasoning Goals)	(b) Mixed-Prior Knowledge Groups (Independent of Goals)

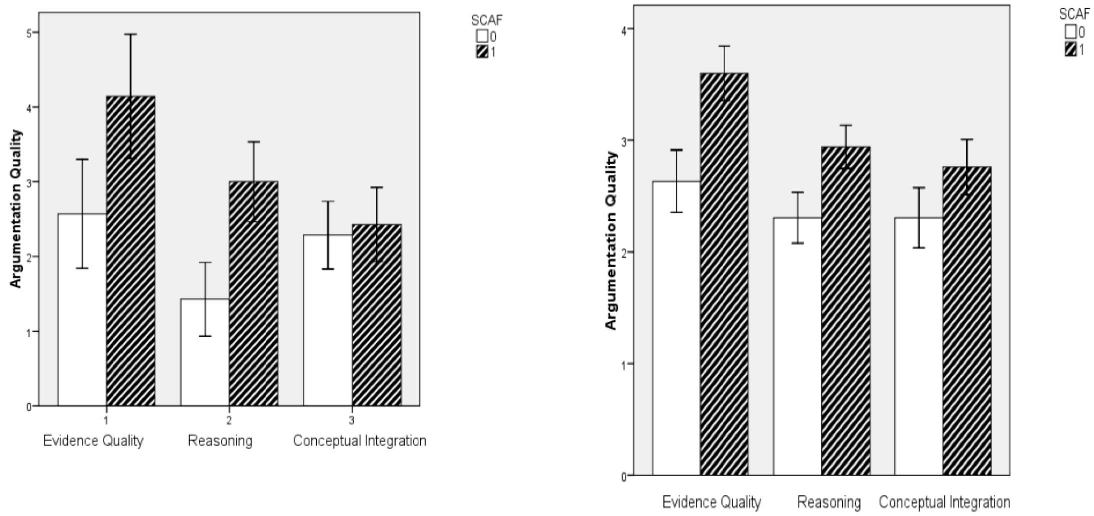

Figure 4.4. Source representation effects across group compositions. SCAF0=checklist only; SCAF1=checklist + annotation. Error bars represent 95% confidence intervals.

As shown in Figure 4.4a, a mean difference was evident in conceptual integration though the effect size was marginal (Cohen's d=.08 SD). The annotation-given groups' adjusted mean score was 2.47 suggesting students were transitioning from the multi-structural to relational level associations among multiple aspects of a given phenomenon.

Mixed-Prior Knowledge Groups. The analysis indicated that only source representation was effective independent of goal instructions ($\Lambda = 0.70$, $F(3,57)=8.06$, $p<.001, \eta_p^2 = .30$). Follow-up ANOVAs indicated statistically significant mean differences between two source representation conditions in terms of evidence quality ($F(1,59)=17.79$, p=<.001, $\eta_p^2=.23$, $d=1.04$) and reasoning ($F(1,59)=14.10$, $p<.001$, $\eta_p^2=.10$, $d=.91$). Groups given the annotation tool performed most effectively. Adjusted mean scores for evidence quality (3.63) and reasoning (2.37) indicated that students were transitioning to consider both converging and alternative evidence and to create warranted arguments. Though not statistically significant ($F(1,59)=2.82$, p=.098, η_p^2=.05), the mean difference of conceptual integration was a potential importance (Cohen's *d*=.41 SD). The annotation groups' adjusted mean score (2.87) suggested that their performance approximated relational levels of conceptual integration.

Transfer of Source Representation and Goal Instructions

We examined individual writing to test potential transfer of scaffolding after removal. As summarized in Table 4.4, both source representation (Λ= 0.98, F(3,451)=3.24, p=.02, η_p^2=.02) and goals (Λ= 0.70, F(3,451)=3.96, p=.008, η_p^2=.03) were statistically significant after controlling for previous individual argumentation scores; however, the effect size explained only 2-3 % of overall performance variance. Two significant interaction effects were also detected—between goals and group compositions and between goals and a covariate (previous reasoning scores)—which limits the generalizability of scaffolding effects. We concluded that little transfer was evident. On average, evidence quality across conditions reached only 2.38 (qualified yet piecemeal evidence); reasoning was 1.87 approximated substantiated arguments without warrants. Conceptual integration (2.9 out of 5.0) did not adequately explain the phenomena.

To examine further the nature of interaction, we examined simple main effects of goal instructions on overall argumentation performance by alternating group compositions and pre-reasoning score. Similarly to the previous results, balanced goals experience benefitted students from low-prior knowledge groups (Λ= 0.98, F(3,464)=2.76, p=.042). Balanced goals also benefitted students with lower (25%) pre-reasoning scores (Λ= 0.97, F(3,451)=3.96, p=.008). The findings together indicated that while transfer effects were marginal they tended to favor students in the low-prior knowledge groups.

Table 4.4

Source Representation and Goal Instruction Effects on Individual Argumentation (N=482)

Factor	Dependent Variable	Conditions	N	Indi. Task1		Indi. Task2		MANOVA			
				M	SD	Adj.M	SD	Λ	F(3,451)	p	η_p^2
Source	Evidence Quality	Checklist	229	2.36	1.38	2.31	.74	.98	3.24	.022	.02
		Annotation+	253	2.52	.95	2.45	.64				
	Reasoning	Checklist	229	2.20	1.00	1.84	.74				
		Annotation+	253	2.39	.54	1.90	.80				
	Conceptual Integration	Checklist	229	3.21	.79	2.93	1.19				
		Annotation+	253	3.02	1.00	2.91	1.27				
Goal	Evidence Quality	Persuasion	249	3.09	.90	2.41	.62	.70	3.96	.008	.03
		Balanced	234	2.88	.98	2.34	.76				
	Reasoning	Persuasion	249	2.62	.78	1.86	.78				
		Balanced	234	2.22	.82	1.88	.76				
	Conceptual Integration	Persuasion	249	1.68	1.04	2.98	1.25				
		Balanced	234	1.75	.92	2.87	1.37				

Note. Evidence quality 0 (no evidence), 1 (unqualified evidence), 2 (qualified evidence), 3 (converging evidence), 4 (alternative evidence), 5(dialectic evidence); Reasoning 0 (unjustified), 1(unsubstantiated), 2(substantiated), 3-4 (warranted); Conceptual integration (analytic score sum) 0 to 5.

Experience Using Scaffolds

Goal Instructions and Task Perceptions. We expected balanced reasoning goal instructions to establish expectations for critical reasoning and higher evaluation standards. Analysis of students' task perception statements (N=482) revealed *how goal instructions influenced task expectations* and subsequent information-seeking and -evaluation efforts in real-time, complex classroom situations. Two-step cluster analyses were conducted to summarize individual perception profiles within each persuasion and balanced reasoning goal condition based on similarity[6] among attributes (refer to Appendix N for the results and descriptive statistics). Table 4.5 describes each cluster, which was defined based on key attributes ($p < .05$ against chi-square distribution) with corresponding student responses.

[6] We computed similarity measures using log-likelihood distance; Schwarz's Bayesian Criterion (BIC) and chi-square statistics determined the optimal number of clusters and key attributes separating clusters (Bacher, Wenzig, & Vogler, 2004; Norusis, 2008).

Table 4.5. Task Perception Profile Description (N=482)

Goals	Cluster (%)	Name	Description	Selected Attributes
Persuasion (N=248)	1 (23)	Inquiry-Dislikers	Consider inquiry as time-consuming and prefer lectures	Learning task (H), No utility (H) Topical interest (H)
	2 (32)	Pseudo-Inquirers for Better Grades	Concern what the instructor is looking for	Fixed science (H), For-grades task (H), Easy evaluation (H)
	3 (22)	Mechanical Evaluators	Appreciate valid sources, but the answer is already fixed.	Open science (**L**), Evaluation task (H), Intrinsic evaluation (H) No prior knowledge (H)
	4 (23)	Knowledge-able Explorers	Expect the answer to be conditional based on their background knowledge and exhibit high reflection.	Open science (H) Reflective evaluation (H) Prior knowledge (H)
	Total (100)			
Balanced Reasoning (N=234)	1 (33)	Inquiry-Dislikers	Consider inquiry as time-consuming and prefer lectures	Learning task (H) No utility (H), difficulty (**L**)
	2 (28)	Biased-Inquirers	Consider inquiry as straightforward and elementary, and search for information to confirm known facts	Fixed Science (H) Research Task (H) For-grades task (H) Prior knowledge (H)
	3 (12)	Trial-and-Error Learners	Conduct initial quick and easy search, which is elaborated over time in interaction with group members or resources, and exhibit high reflection.	Easy evaluation (H) Reflective evaluation (H) No prior knowledge (H)
	4 (14)	Reflective Explorers	Recognize openness in science inquiry though their evaluation strategies are not clear.	Open science (H) Reflective evaluation (H)
	5 (13)	Active Evaluators	Appreciate inquiry and information evaluation; exhibit intrinsic, extrinsic, reflective evaluation.	Evaluation task (H) Intrinsic evaluation (H) Extrinsic evaluation (H) Topical Interest (H)
	Total (100)			

Note. H and L mean high and low % frequency respectively (See Appendix C for the attribute definition)

Students given persuasion goals (control; N=248) composed four clusters. The clusters reflected low or mechanical evaluation standards and efforts (c.f., Tsai, 2004; Wu & Tsai, 2005) with the exception of *Knowledgeable Explorers* (23%) who identified the controversial nature of the inquiry topic based on their background. Despite interest in given topics, *Inquiry-Dislikers* (23%), for example, considered inquiry as "frustrating" and they "liked to have biology rather than method." They noted varying evaluation strategies, which however failed to influence their information-seeking efforts: "the amount of effort I was expected to put …seemed waste of time … we were only expected to submit two pages of actual information." *Pseudo-Inquirers for Better Grades* (32%) believed that a correct answer was available "the instructor was looking for" and focused on fewer "proven" and presumably "persuasive" sources (e.g., New York Times). *Mechanical Evaluators* (23%) commented that valid sources are key to "narrow down information." Whereas referring to scholarly journals because "it is the correct answer," or "it sounds persuasive," they did not incorporate much information due to technical jargon or complex conceptual structures, which expert readers often resolve using extrinsic evaluation or comparing similarities and distinctions across sources (c.f., Wiley et al., 2009).

Students given balanced reasoning goals (N=234) composed five clusters indicating segmentation in task perceptions. The majority of students were classified as *Inquiry-Dislikers* (33%) or *Biased Inquirers* (28%). Typically, they ignored or did not focus on explicit directions to "consider counter-arguments and explain why they were wrong" because "(counter-arguments) would make things complicate." For example, *Biased Inquirers* were considered familiar with given inquiry topics. While conceding the importance of counter-evidence, they were overconfident about their position—"it was way straightforward."

The remaining clusters demonstrated unexpected, transitional, or critical evaluation standards, which often involved positive conceptual or beliefs change during inquiry. *Trial-and-Error Learners* (12%) typically possessed limited background knowledge; initial information-seeking and -evaluation involved little elaboration since "(they) did not know where to start." Students described their learning rather as "trial-and-error" or "enlightening." 93% *Trial-and Error Learners* indicated key contributors were dual-position documents where "the author covered both sides…much easier to understand" or their group-work. Students "compared (their) findings, (were) surprised, but enlightened, to find it took much more digging and balancing (between different ideas)." *Reflective Explorers* (14%) comprised diverse topic interest and perceived difficulty. Although they did not articulate intrinsic or extrinsic evaluation strategies, most expected their inquiry to be complex or controversial after reading the task direction. They reported changes in their claims or beliefs after continual information re-seeking and organization (reflective evaluation). Finally, Active Evaluators (12.9%) acknowledged the importance of evaluation and articulated their evaluation strategies including intrinsic, extrinsic, and reflective evaluation. These findings together indicate that while goal instructions seem to serve as triggers for evaluation standards when combined with other conditions (e.g., prior knowledge levels, interest, group works, and ancillary documents), goal instructions alone may prove insufficient to induce higher evaluation standards in complex classroom situations (c.f., Nussbaum & Kardash, 2005).

Tool Use and Evaluation. We expected synergistic effects of source representation combined with balanced goals (Tabak, 2004) on evaluation and argumentation. To examine tool use experience (*how tools did or did not benefit students)*, an open-response survey item queried students to describe their annotation tool or checklist use experience. The thematic analysis of student responses (N=482) identified tool properties and common use cases—cue recognition, filtering, articulation, and claim strengthening. Categories consist of dimensions differentiating tool use experience between the checklist and annotation tools. We counted dimension frequencies within each tool respondents; when needed, dimensions have been refined again into sub-dimensions. Table 4.6 summarizes each category and its dimensions differentiating tool use experience.

Table 4.6. Tool Properties and Use Cases Category

Category	Sub-Category	Definition	Dimensions
Tool Properties	• Communication Mode	-Descriptions about tool interaction methods	-Verbal Prompts (Acronym) vs. Non-verbal Constraints (Visualizer)
	• Specificity	-Descriptions of domain or task specific cues that the tool supports	-Generic vs. Task specific
	• Coverage	-Descriptions of evaluation support areas	-Predictive, Intrinsic, Extrinsic, Reflective Evaluation Cues
	• Cognitive Dependency	-Descriptions about division of labor between the user and tools	-Stand-alone vs. Distributed (e.g., annotation as a memory aid/reference)
Use Cases	• Cue-Recognition	-Tool use to increase awareness of source properties	- Before, During, After Search
	• Filtering	-Tool use to narrow down the amount and scope of sources	-Efficiency Centered vs. Claim-Centered
	• Articulation	-Tool use to communication source properties	-Individual vs. Relational - Mechanical/Momentary vs. Elaborate/ Continual
	• Claim Strengthening	-Tool use to link evidence to the claim	- Validation /Confidence vs. Balancing /Understanding -Individual vs. Group Reasoning

The checklist incorporated acronym-based prompts (CRAP), and listed generic evaluation criteria (currency, reliability, authority, and purpose) and exemplary cues. The annotation tool visualized and scaffolded source evaluation steps specific to the argumentation process. 45% of checklist users (control; N=248) noted that the CRAP acronym was easy to retrieve during other cognitive activities (e.g., search). However, the system use resulted in only limited or mechanical source representations; 87% of comments on tool affordances focused on intrinsic evaluation. Students simplified sources to “primary” or “scholarly” sources. The majority of annotation tool students (treatment; N=234), in contrast, commented source compiling (85%) and claim-evidence linking (81%) functions were beneficial. This memory aid and reference enabled continual "claim-prompted" evaluation throughout the task.

Students under both control and treatment conditions reported that tool use increased cue-recognition, filtered sources, helped to articulate reasoning behind evaluation, and strengthened claims; however, the influence varied across conditions. The checklist was used mainly during search/scanning (65%) and reference writing process (85%). Checklist users narrowed down sources and increased "efficiency" during initial inquiry, which limited conceptual elaboration across diverse claims. While they articulated properties for sourcing or write-up, 87% of those properties were limited to innate qualities of individual sources. 35% students commented their reasoning often did not vary—"I often felt that was repeated and redundant." Finally, evaluation was considered as a minimum requirement for the task completion. Only two checklist users mentioned that reliability criteria helped to compare sources and understand the topic; group negotiation for sourcing and sensing was not reported. Most students (62%) merely reported that evaluation increased their confidence about the claim "early enough" to complete the task efficiently. Other students (30%) complained that efforts to improve efficiency limited time to access otherwise usable evidence—"sometimes it is too 'simple' when used to figure out a source. Sometimes an 'old' source can actually be a really good one."

The annotation tool use was widely used across search (20%), reading (32%) organizing (40%), group discussion (33%), and (re-) writing (80%). Students filtered sources to "examine" and "balance" their claims through continual evaluation: "we used it to narrow our search down to source of information that best fit our topic, as well as counter-evidence to show all perspectives." They articulated source properties in relation to claims and source relations. Comments on tool affordances focused on extrinsic evaluation (95%), reflective evaluation

(60%), and predictive evaluation (30%) for more information-seeking and -elaboration. 65% of the annotation users commented that the tool helped to track "what information (they) had and what (they) needed in the tool;" supported re-evaluation and organization of "thoughts and sources;" and helped to "reach a better conclusion." The tool also served as shared reference to negotiate among group members (32%): "the most helpful aspect of the tool is that while our group could individually research, the tool was a way to bring all our work to the same place and was an organized layout."

Tool Use Cognitive Load. Students also reported the influence of increased task steps on the cognitive load (N=464). NASA-Task Load Index analysis across source representation conditions confirmed this observation. As shown in Table 4.7, the results indicated a significant difference between annotation and checklist use, $t(462) = 3.87$, $p < .001$ indicating that annotation increased cognitive load (Cohen's *d*=.35 SD).

Table 4.7

NASA-Task Load Index (TLX) Score Analysis (N=464)

Sub-Scale	Condition	M	SD	Index Score				t-test		
				Condition	N	M	SD	t(462)	p	*d*
Mental Demand	Checklist	70.67	23.47	Checklist	229	271.83	78.59	3.87	.00	.35
	Annotation	70.10	23.43							
Time Pressure	Checklist	63.93	26.39							
	Annotation	72.52	23.78							
Efforts	Checklist	77.91	21.46	Annotation	235	297.62	64.32			
	Annotation	76.45	20.07							
Frustration	Checklist	59.32	30.66							
	Annotation	78.55	23.14							

Note. Possible index score ranged from 0 to 400.

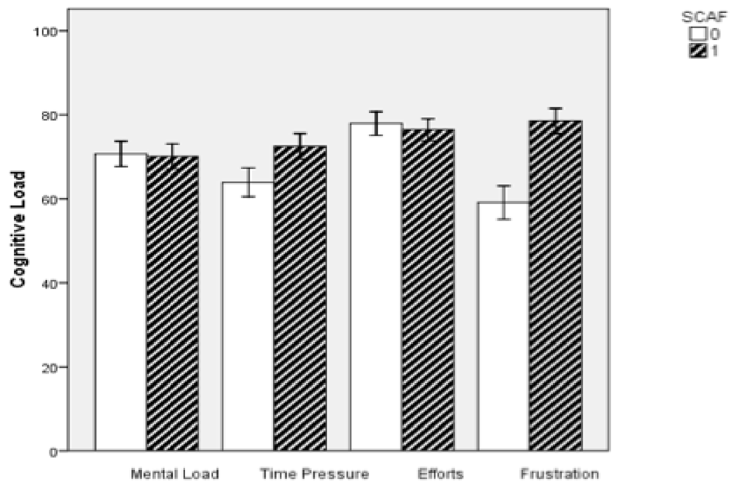

Figure 4.5. NASA-Task Load Index sub-scale means. SCAF0=checklist only; SCAF1=checklist + annotation. Error bars represent 95% confidence intervals.

In order to identify the source of load, we examined sub-scale mean differences between conditions. As Figure 4.5 illustrates, task demands and associated effort were comparable suggesting that perceived difficulty and cognitive effort were not significantly greater for the annotation tool (c.f., Gerjets, Kammerer, & Werner, 2011). However, use of the annotation tool involved increased time-pressure and frustration.

Student responses to tool use also provided insights to interpretation of the results: 35% of students reported the tool was not difficult to use whereas 22% commented there were "too many steps simultaneously going on." Since the annotation involved continual micro-evaluative steps, the increased steps, though important for learning, might have competed for limited cognitive resources and resulted in frustration. Others (27%) noted the tool requirements were important but unnecessary for college students given time allotted. Though such comments may reflect overconfidence (c.f., Winne, 2004), negative task perceptions and motivational beliefs might cause pseudo-redundancy effects (c.f., Kalyuga, 2007) where learners consider otherwise essential instructional features to be extraneous. Still other students (25%) reported difficulties transferring from search activities to the tool use and writing, or from individual to group activities. For example, students documented source properties (e.g., author names and URLs) in the tool after individual search and then included them in their group write-up. The tool thus may have involved repeated, unnecessary steps, which increased student frustration (c.f., Zhang & Quintana, 2012). These findings suggested that perceived cognitive load here might need to be weighed with negative motivational beliefs and perceptions of simultaneous or extraneous steps that may overtax limited cognitive capacity.

General Discussion

Information evaluation is a critical yet rarely activated successfully for academic tasks (e.g., argumentation). In a previous study, we reported that source representation and goal instructions increased first-year college students' predictive, intrinsic, extrinsic, and reflective evaluation (Kim & Hannafin, 2014). Since the study findings were based largely on self-reports, the current study examined whether and how treatment conditions influenced students' actual evaluation skill use to influence argumentation quality.

Goal Instructions in the Complex Classroom Situations

As different from given-source evaluation conditions, online inquiry does not ensure that alternative sources will be identified or examined for evaluation, which are considered pre-requisite to critical argumentation. Unless students intentionally seek alternative sources to verify existing sources, they are apt to naively consider incomplete resources. To induce proactive, or positive predictive evaluation, we used goal instructions (Ferretti et al., 2000; Nussbaum & Kardash, 2005; Page-Voth & Graham, 1999).

However, despite positive potential to support low-prior knowledge students using an annotation tool, balanced goals did not demonstrate independent effects on group and individual argumentation. The findings suggest balanced goals alone were insufficient to induce higher evaluation standards and efforts in complex classroom situations (c.f., Nussbaum & Kardash, 2005). The written direction rather served as a weak trigger which induced negative or positive evaluation standards depending on other classroom conditions (e.g., perceived values, prior knowledge levels, ancillary documents, and group works). The direction ("consider counter-arguments and explain why they were wrong") was either devalued or unacknowledged for some students but proved prominent for others; some students acknowledged the direction only after interaction with group members or other documents that conveyed diverse views. The trends of individual writing scores confirmed this observation. Though the effect size was marginal, balanced goals benefited low-prior knowledge student groups and those with low pre-reasoning scores. While investigations of these moderating factors were beyond the scope of this study, this finding reinforces cautions that designers' intended goals do not necessarily align with students' perceived and activated goals. It is imperative to consider students' perceived goals to enhance regulatory learning in information-rich environments (Manlove, Lazonder, & Jong, 2009; Winne & Nesbit, 2009).

Findings and guidance from goal and metacognition researchers may require several considerations. First, goal strength of direction may enhance the visibility and utility of goal instructions (Kaplan & Rothkopf, 1974; Locke & Latham, 2002). For example, *Inquiry Dislikers* and *Biased-Inquirers* in the current study devalued or ignored given goals based on their individual history, preferences (e.g., quick learning) or overconfidence. Specifying why the goal is important (i.e., metacognitive knowledge of strategy utility) may help students to recognize the goal as intended and have a strong impact on learning (Borkowski, Carr, Rellinger, & Pressley, 1990; Jiang & Elen, 2011; Oliver & Hannafin, 2000). A related consideration involves goal negotiation (Stahl, 2006; Winne & Nesbit, 2009), considered important process to aligning instructional goals and perceived goals in classroom (Song, Hannafin, & Hill, 2007). This aligning or reconciliation process may be mediated by group dynamics (Reason, 1999), feedback (Locke & Latham, 2002), or instructional objects (e.g. dual-position texts or cases) (Nussbaum & Kardash, 2005) that outline alternative arguments. In the current study, *Trial-and-Error Learners* were reminded to consider balanced goal instructions during group interactions and advanced group members to weigh alternative information sources or perspectives. Lastly, it is important to activate and sustain goals throughout to support evaluation and argumentation. When given regulatory aids (e.g., source representation tool), students tend to activate their perceived goals to perform the task (Winne & Nesbit, 2009).

Source Representation and Cognitive Tool

The annotation tool modeled advanced cues and standards by integrating argumentation structure and evaluation; the tool also served as a memory aid that registered multiple representations to free one's cognitive capacity. Thus we expected students who used the annotation tool to attend more intently to extrinsic and reflective cues, tag cues to organize information, and deepen more conceptual integration. This assumption was partially supported. Despite an impressive effect size explaining 30 to 49% of performance variability, the annotation was only effective when given balanced reasoning goals or in heterogeneous knowledge groups. Also, students reported germane as well as extraneous cognitive load, which required re-examination of scaffolds design principles.

First, the annotation tool supported groups comprising mixed-prior knowledge to enhance evidence quality and reasoning though the effect size for conceptual integration was moderate. However, homogeneous low-prior knowledge groups performed better only when the annotation tool and balanced reasoning goals were combined. The finding suggests a synergy between source representation and goal instructions. Students might not fully use the tool without proper tasks representation (i.e., requirements) and accordingly calibrated standards but tend to use more effectively when balanced reasoning goals are scaffolded. This may also apply to mixed-prior knowledge groups, where more capable members might proxy for the balanced reasoning goal instructions role (Belland, 2010; Carter & Jones, 1994; White & Frederiksen, 1998) thus minimizing the effect of goal instructions. Our follow-up analysis of students' experiences support this influence. Though the tool helped to externalize and integrate 'claim-prompted' evaluation, negative task perceptions and beliefs limited its use. Given synergistic support, both low and mixed-prior knowledge groups transitioned to include alternative evidence, warranted arguments, and relational level of conceptual understanding. These represent important potential improvements considering college students' confirmation bias, unsubstantiated or unwarranted explanation (Brem & Rips, 2000; Kuhn, 1989), and lack of spontaneous elaboration (Asher, 2011).

However, no transfer effect was evident for individual argumentation after the intervention support was removed. First, there might been free-rider effects, which are the frequently reported problems in group learning (Karau & Williams, 1993; Slavin, 1995). However, in our previous study, we noted that annotation effects for self-reported information evaluation behavior increased over time, suggesting transfer in term of metacognitive awareness. Alternatively, the annotation tool's compiled sources properties might serve as a memory and reference for review; removal, therefore, might have decreased productivity and monitoring effects. Also, decreased motivation of task completion rate (i.e., 84% completion), might influence the individual's writing quality. Unfortunately, the current study did not test these hypothesis. A further study is needed.

Finally, it is important to analyze how students actually used the annotation tool and checklist. Though simple and efficient, the checklist encourages mechanical search and depth-first search, or tendency to fix the information search scope early to increase depth rather than breadth of search (c.f., Guo, 2011). Reliability, for example, should be judged based on a specific claim-evidence structure (c.f., Brem, Russell, & Weems, 2001) and evidence comparisons (c.f., Meola, 2004). Students, however, simplified reliable sources to "primary" or "scholarly" sources. Also, students noted that the checklist increased confidence about claims early on; however, depth-first search is often associated with back-chaining reasoning, naïve conclusion, and superficial understanding (Kirschenbaum, 1992). Thus, cautions should be exercised against mechanical uses of imposed functionality (Sharma & Hannafin, 2007; Zimmerman & Tsikalas, 2005). For example, students in the current study perceived that the requirements for reference-section write up involved seemingly repetitive, time-consuming reasoning. The can be modified to require students to rank or contrast source quality in light of their claims.

Conversely, the annotation tool balance depth-first search with breadth-first search. Students used the CRAP criteria to narrow down the main points, but the tool forced students to consider alternative ideas, engage in further information seeking, and develop elaborated explanation to reach a conclusion. Students cited the visualized, claim-centered structure and reference for memory, reflection, and group negotiation as appreciated. However, tool use involves the multiple, integrated cognitive tasks of search, scanning, reading, and integration. Since these phases are rapidly changing and often processed in a parallel manner (c.f., Bowler, 2010), the tool may compete for cognitive capacity during operation and cause confusions in the uses of different tools (e.g., browser and word-processor) causing cognitive load and frustration. Jonassen (2010) suggested the need for increased attention to the timing and placement of metacognitive scaffolding. Multiple tools integration (Lee, 2004; Miyata & Norman, 1986; Zhang & Quintana, 2012) and mechanical activities automation should help students to concentrate on important cognitive activities (Kim & Reeves, 2007; Quintana, Zhang, & Krajcik, 2005; Zhang & Quintana, 2012). Promising alternatives may emerge as browsers and search engines integrate main cognitive activities with critical metacognitive reflection. For example, Ennals, Trushkowsky, and Agosta (2010) designed Dispute Finder, a browser extension that directs users to both converging and counter-evidence sources to develop a critical argument without splitting attention. Zhang and Quintana's (2012) Digital IdeaKeeper automatically saves log files to free students investing effort to to save and store search results during inquiry.

Conclusion

Our findings related to students' use of source representation scaffolds and goal instructions during inquiries provide evidence that the effective argumentation-based evaluation results from the synergistic integration of scaffolding sources. In metacognitive terms, information evaluation involves continually defining a task, registering, interpreting, and comparing cues with standards and interpreting differences to warrant appropriate actions or judgments (Pressley, 2002; Rouet, 2006; Schraw, 1998; Winne & Nesbit, 2009). By addressing student perceptions of a given task (balanced reasoning goals), increasing cue recognition, and providing a memory aid (annotation tool) to compensate for cognitive capacity, scaffolding spanned the processes to promote self-sustaining, metacognitive learning (c.f., Zimmerman & Tsikalas, 2005). However, the feasibility if providing multiple additional scaffolds to support metacognitive activities (e.g., information evaluation) in large-enrollment college classrooms remains a complex issue. Scaffold design and research needs to consider several factors (e.g., prior knowledge levels, group dynamics, perceived values, and knowledge of strategy utility) to ensure that students recognize the intended goals, use and sustain the intended the functionality.

References

Asher, A. D. (2011). Search magic: Discovering how undergraduates find information. Presented at the American Anthropological Association Annual Meeting, Montreal, Canada.

Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS twostep cluster: A first evaluation*. Lehrstuhl für Soziologie. Retrieved from http://www.statisticalinnovations.com/products/twostep.pdf

Belland, B. R. (2010). Portraits of middle school students constructing evidence-based arguments during problem-based learning: the impact of computer-based scaffolds. *Educational Technology Research & Development*, *58*(3), 285–309. doi:10.1007/s11423-009-9139-4

Belland, B. R., Glazewski, K., & Richardson, J. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research & Development*, *56*(4), 401–422. doi:10.1007/s11423-007-9074-1

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning : the SOLO taxonomy (structure of the observed learning outcome)*. New York : Academic Press, 1982.

Borkowski, J. G., Carr, M., Rellinger, E., & Pressley, M. (1990). Self-regulated cognition: Interdependence of metacognition, attributions, and self-esteem. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 53–92). Hillsdale, NJ: Erlbaum.

Bowler, L. (2010). A taxonomy of adolescent metacognitive knowledge during the information search process. *Library & Information Science Research*, *32*(1), 27–42. doi:16/j.lisr.2009.09.005

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science*, *24*, 573–604. doi:10.1016/S0364-0213(00)00033-1

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition & Instruction*, *20*(4), 485.

Carter, G., & Jones, M. G. (1994). Relationship between ability-paired interactions and the development of fifth graders' concepts of balance. *Journal of Research in Science Teaching*, *31*(8), 847–856.

Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27–43. doi:10.1016/0959-4752(94)90017-5

Cho, K.-L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, *50*(3), 5–22.

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, *45*(3), 293–321. doi:10.1002/tea.20216

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010). Highlighting disputed claims on the Web. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 341–350). New York, NY, USA: ACM. doi:10.1145/1772690.1772726

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ (Clinical Research Ed.)*, *324*(7337), 573–577.

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, *92*(4), 694–702. doi:10.1037/0022-0663.92.4.694

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. doi:doi: 10.1016/j.learninstruc.2010.02.005

Guo, B. (2011). The scope of external information-seeking under uncertainty: An individual-level study. *International Journal of Information Management*, *31*(2), 137–148. doi:10.1016/j.ijinfomgt.2010.08.005

Hae-Deok Song. (2005). Motivating ill-structured problem solving in a Web-based peer-group learning environment: : A learning-goal perspective. *Journal of Educational Computing Research*, *33*(4), 351–367.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (Vol. Volume 52, pp.

139–183). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623869

Head, A. (2013). Project information literacy: What can be learned about the information-seeking behaviour of today's college students? In *Association of College and Research Libraries (ACRL) Proceedings 2013*. Chicago: ALA.

Head, A., & Eisenberg, M. (2009). Lessons learned: How college students seek information in the digital age. Retrieved from http://projectinfolit.org/pdfs/PIL_Fall2009_finalv_YR1_12_2009v2.pdf

Hirsch, R. (1989). *Argumentation, information, and interaction: Studies in face-to-face interactive argumentation under differing turn-taking conditions*. Department of Linguistics, University of Göteborg.

Hjørland, B. (2012). Methods for evaluating information sources: An annotated catalogue. *Journal of Information Science*, *38*(3), 258–268. doi:10.1177/0165551512439178

Howell, G. T., & Lacroix, G. L. (2012). Decomposing interactions using GLM in combination with the COMPARE, LMATRIX and MMATRIX subcommands in SPSS. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 1–22.

Hughes, B., Wareham, J., & Joshi, I. (2010). Doctors' online information needs, cognitive search strategies, and judgments of information quality and cognitive authority: How predictive judgments introduce bias into cognitive search models. *Journal of the American Society for Information Science & Technology*, *61*(3), 433–452. doi:10.1002/asi.21245

Jiang, L., & Elen, J. (2011). Why do learning goals (not) work: a reexamination of the hypothesized effectiveness of learning goals based on students' behaviour and cognitive processes. *Educational Technology Research and Development*, *59*(4), 553–573. doi:10.1007/s11423-011-9200-y

Jonassen, D., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development*, *58*(4), 439–57. doi:10.1007/s11423-009-9143-8.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. doi:10.1007/s10648-007-9054-3

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410. doi:16/j.learninstruc.2009.02.003

Kalyuga, S. (2011). Informing: A cognitive load perspective. *Informing Science*, *14*, 33–45.

Kaplan, R., & Rothkopf, E. Z. (1974). Instructional objectives as directions to learners: Effect of passage length and amount of objective-relevant content. *Journal of Educational Psychology*, *66*(3), 448.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681.

Kim, B., & Reeves, T. (2007). Reframing research on learning with technology: In search of the meaning of cognitive tools.

Kim, S., & Hannafin, M. (2013). Science on the Web: How do college students evaluate and support arguments? Manuscript submitted for publication.

Kim, S., & Hannafin, M. (2014). The effects of source representations and goal instructions on college students' information evaluation behavior change. Manuscript submitted for publication.

Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology*, *77*(3), 343–352. doi:10.1037/0021-9010.77.3.343

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689. doi:10.1037/0033-295X.96.4.674

Lee, Y.-J. (2004). Concept mapping your Web searches: a design rationale and Web-enabled application. *Journal of Computer Assisted Learning*, *20*(2), 103–113. doi:10.1111/j.1365-2729.2004.00070.x

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. doi:10.1037/0003-066X.57.9.705

Manlove, S., Lazonder, A. W., & Jong, T. de. (2009). Trends and issues of regulative support use during inquiry learning: Patterns from three studies. *Computers in Human Behavior*, *25*(4), 795–803. doi:10.1016/j.chb.2008.07.010

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Commun. ACM*, *49*(4), 41–46. doi:10.1145/1121949.1121979

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy*, *4*(3), 331–344.

Miyata, Y., & Norman, D. A. (1986). The control of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design : new perspectives on human-computer interaction* (pp. 265–284). Hillsdale, N.J.: L. Erlbaum Associates.

Norusis, M. (2008). *SPSS 16.0 Statistical Procedures Companion* (2 edition.). Upper Saddle River, NJ: Prentice Hall.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. doi:10.1037/0022-0663.97.2.157

Oh, S., & Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning*, *23*(2), 95–110. doi:10.1111/j.1365-2729.2006.00206.x

Oliver, K., & Hannafin, M. J. (2000). Student Management of Web-Based Hypermedia Resources During Open-Ended Problem Solving. *Journal of Educational Research*, *94*(2), 75.

Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, *91*(2), 230–240. doi:10.1037/0022-0663.91.2.230

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What Research Has to Say About Reading Instruction* (pp. 291–309). Newark, DE: International Reading Association.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A Framework for supporting metacognitive aspects of online inquiry yhrough software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. doi:10.1207/s15326985ep4004_5

Reason, P. (1999). General medical and complementary practitioners working together: The Epistemological Demands of Collaboration. *The Journal of Applied Behavioral Science*, *35*(1), 71–86.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.

Rieh, S. Y., & Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 49–71). Cambridge, Mass. : MIT Press, c2008.

Rouet, J. (2006). *The skills of document use: From text comprehension to Web-based learning*. L. Erlbaum Associates.

Rouet, J.-F. (2009). Managing cognitive load during document-based learning. *Learning and Instruction*, *19*(5), 445–450. doi:16/j.learninstruc.2009.02.007

Russell, D. M., Stefik, M. J., Pirolli, P. L., & Card, S. K. (1993). The cost structure of sensemaking (pp. 269–276). Presented at the ACM SIGCHI conference on human factors in computing systems.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*(1-2), 113–25.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, *15*(1), 27. doi:10.1080/10494820600996972

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (Vol. 2). Allyn and Bacon Boston. Retrieved from http://www.getcited.org/pub/103185225

Song, L., Hannafin, M. J., & Hill, J. R. (2007). Reconciling beliefs and practices in teaching and learning. *Educational Technology Research and Development*, *55*(1), 27–50.

Stadtler, M., Scharrer, L., & Bromme, R. (2011). How reading goals and rhetorical signals influence recipients' recognition of intertexual conflicts. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Sceicne Society* (pp. 1346–1351). Austin, TX: Cognitive Sciecne Society.

Stahl, E., StephanieBromme, Rainer. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research*, *35*(4), 319–338.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, *13*(3), 305–335.

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research (07408188)*, *27*(2), 222–248. doi:10.1016/j.lisr.2005.01.005

Teddlie, C., & Tasshakori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, *13*(1), 12–28.

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

Tsai, C.-C. (2004). Information commitments in Web-based learning environments. *Innovations in Education and Teaching International*, *41*(1), 105–112.

Walraven, A., Brand-gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, *24*(3), 623–648. doi:doi: 10.1016/j.chb.2007.01.030

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346. doi:10.1016/S0364-0213(99)80043-3

Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research*, *41*(6), 466–488. doi:16/j.ijer.2005.08.012

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York: Routledge.

Wu, Y.-T., & Tsai, C.-C. (2005). Information commitments: Evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, *21*(5), 374–385. doi:10.1111/j.1365-2729.2005.00144.x

Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, *58*(1), 181–196.

Zimmerman, B. J., & Tsikalas, K. E. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning? *Educational Psychologist*, *40*(4), 267–271. doi:10.1207/s15326985ep4004_8

CHAPTER 5

CONCLUSION

Information evaluation is a necessary but often unsupported skill when teaching argumentation (Brem, Russell, & Weems, 2001; Iding, Crosby, Auernheimer, & Barbara, 2008; Wiley et al., 2009). This dissertation aimed to understand and support mental tasks and conditions for seasoned information evaluation during argumentation based inquiry. Source representation scaffolds and goal instructions were devised and implemented to test the effects in the college science classrooms. Source representation scaffolds contrasted the annotation tool and the checklist. Goal instructions contrasted balanced reasoing goals and persuasion goals. Given the widespread use of checklists and persuasion goals in science classes, their combination served as a baseline. Chapter 2 discussed theoretical underpinnings which guided scaffolding design and research. Chapter 3 and 4 presented empirical study findings conducted to examine the effects of scaffolds.

Scaffolding Effects on Information Evaluation Behavior

The main research study 1 (chapter 3) examined the scaffolds' effects on information evaluation behavior change while controlling for students' pecieved task values and cognitive capacity. Information evaluation behavior (IEB) self-measures were colleted and analyzed using multilevel analyses. The findings indicated poor initial information evaluation behavior, which was influenced by perceived task values and cognitive capacity; student behavior slowly improved under a baseline condition. On the other hand, both online annotation and balanced reasoning goals improved information evaluation behavior, yet indicated a negative interaction effect reducing the possible combined effects. The findings together supported the prediction that scaffolds improved information evaluation behavior and associated individual conditions. The negative interaction, however, suggested possible redundancy between two scaffolding approaches.

Scaffolding Effects on Argumentation Quality

Since study 1 involved self-reports, which suggested metacognitive awareness, research study 2 (chapter 4) examined whether and how treatment conditions influenced students' actual evaluation skill use to influence argumentation quality. Group and individual writing samples were examined during and following scaffolding interventions. Scaffolds experience and cognitive load were examined. The results were mixed. Balanced reasoning goal instructions did not directly influence argumentation quality. Students' task perceptions suggested considering goal strength, goal negotiation, and goal regulation. Conversely, the annotation tool (as a regulatory, modeling, and memory aid) had significant effects but only when the annotation tool was combined with balanced reasoning goals or was implemented by heterogeneous prior knowledge groups. There were no transfer effects possibly due to the removal of a regulatory and memory aid. The findings together supported the synergistic integration of two scaffolding functionality, yet suggested addressing possible difficulties in using scaffolds and sustaining the effects in the complex classroom situations.

Implications for Scaffold Design and Implementation

To appropriately support students, it is necessary to recognize that information evaluation requires complex metacognitive endeavors involving task interpretation, goal setting, multiple source representations with differing focuses, and thus working memory capacity for processing. Accordingly, support needs to be provided throughout the evaluation process to promote self-sustaining metacognitive learning (c.f., Zimmerman & Tsikalas, 2005), as well as to address related mental conditions (e.g., task perceptions and cognitive capacity). Without a proper task representation, calibrated goals, and support for cognitive capacity, seasoned information evaluation would not easily occur.

In short, this program of research demonstrated the potential of synergistic, though often redundant, integration of goal instructions and source representation tools. The balanced reasoning goals and annotation tool served the same functionality of increasing cue recognition, and thus their combination did not reach the expected extent of benefits (i.e., double effects). However, it might have created multiple opportunities for everyone, especially low performing students, to interact with the intended functionality. The balanced reasoning goals and annotation tool, on the other hand, augmented each other in terms of triggering and regulating the need for critical information evaluation. Thus, each scaffold given the other could reach the intended goal of scaffolding functionality.

However, cautions against mechanical uses of multiple, repeatedly imposed functionality, which might hamper the intended productive synergy, must be heeded (c.f., Tabak, 2004). Each scaffold should be visible and meaningful enough yet relevant to each other sharing the same language and structure across multiple sessions of events. To ensure concerted efforts to create the intended synergy, coherence should be ensured between scaffolds, as well as between activities, and between the designer and instructor. In addition, the beneficial functionality could prove unnecessary and redundant for advanced learners. It should be considered to negotiate the meaning, or to differentiate or reorder the same activities for advanced learners.

Finally, additional scaffolds to support information evaluation should not compete with important cognitive activities for cognitive capacity, as is often the case for other metacognitive or higher-order cognitive activities (Sharma & Hannafin, 2007; Zimmerman & Tsikalas, 2005). Multiple tools integration (Lee, 2004; Miyata & Norman, 1986; Zhang & Quintana, 2012) and mechanical activities automation should help students to concentrate on important cognitive activities without split attention (Kim & Reeves, 2007; Quintana, Zhang, & Krajcik, 2005; Zhang & Quintana, 2012). Promising alternatives may emerge as browsers and search engines integrate main cognitive activities with critical metacognitive reflection. For example, Ennals, Trushkowsky, and Agosta (2010) designed Dispute Finder, a browser extension that directs users to both converging and counter-evidence sources to develop a critical argument without splitting attention. Devising easier approaches the instructor can incorporate to the large, complex classrooms is another remaining task.

Future Research Directions

Both formal science education and everyday science literacy rely increasingly on the Internet for information. Though seemingly simple and taken-for-granted, information evaluation is a critical, complex, yet less activated component of online inquiry and argumentation. This dissertation aimed to understand and support mental tasks and conditions for seasoned information evaluation during argumentation based inquiry. The findings suggest information evaluation is rather dynamic, complex interaction among students, information, and tasks (c.f., Kuiper et al., 2005). Considering prevalent online inquiries and evaluation tasks in everyday and academic tasks, further research should enhance our understanding of the dynamics.

First of all, enhancement of information evaluation during argumentation requires a clearer picture of the associated process and better understanding of current student capacity. This study, building on both direct and indirect results, has proposed a hypothetical reference model of information evaluation. Though some evidence (e.g., students' scaffold use experience) supported that our model is close to reflect what happens during inquiry-based argumentation, the main focus of empirical studies was rather to examine the assumed benefits of devised scaffolds. More empirical studies that shed light on the argumentation process and student difficulties with information evaluation are needed. Unobtrusive observation of student inquiry may serve this purpose. Related to this agenda, contrastive case studies of a baseline condition, checklist use, and annotation use are interesting and meaningful endeavors considering different strength and weakness of tool affordances.

Further examination of individual variables is also needed to design, refine, and sustain the effects of scaffolds. Out findings suggested individual differences explained the initial status as well as the growth of information evaluation behavior. For example, researchers suggest the tendency toward deep thinking (i.e., need for cognition) or epistemic beliefs influence information-seeking (e.g., Mason, Boldrin, & Ariasi, 2009). If so, these conclusions may well be linked to information evaluation behavior, as well as different use of provided scaffolds.

More importantly, however, information evaluation includes dynamic interaction among students, information, and task variables (Kuiper et al., 2005; Lazonder & Rouet, 2008; Rieh & Hilligoss, 2008). The existing approaches to information evaluation instruction often ignore such dynamics in the context of real tasks (Wolfe, 2008). Future studies need to examine the combined effects of different variables. For instance, student characteristics (e.g., levels of prior domain and discourse knowledge), task requirements (e.g., fact finding, summary, and argumentation; individual and collaborative tasks), and the nature of information (e.g., pre-selected information, free-searched information, or both) could bring about different interaction effects. Future studies investigating these effects might provide better rationales for various scaffolding strategies design and maintenance.

Related to this agenda, this dissertation has relatively less focus on differing inquiry topics across tasks. Though distribution of topics for assigned interventions was homogeneous and thus should have less impact on the current studies, it is readily assumed different topics (e.g., complex biological mechanism vs. controversial socio-scientific issues) may incur different approaches to information seeking and evaluation. Inquiries towards to topical influences are another interesting endeavor.

Finally, the Internet space and media rapidly diversifies into social media and user-generated contents. Considering the different structure and sources of information, research needs to be extended to examine information seeking and evaluation behavior on the new media. For example, on another project, the researcher has examined the community structure and information sources of informal health science learning networks on Twitter. Differently from the assumed benefits, Twitter created closed community structure and limited the information flow. It is readily assumed information evaluation standards are naïve and relaxed for this informal network; biased information is detrimental for reasoned, every-day decision making.

Limitations

Certain limitations should be acknowledged. Among others, it should be noted that argumentation has been defined from information seeking perspectives. Though argumentation, rigorously speaking, entails assumption check, grounded reasoning, and explanation building, the current dissertation study significantly focused on evidence linking aspects of argumentation. Approaches to relate information evaluation to other aspects of argumentation (e.g., reflective evaluation and assumption check) are needed.

Other practical limitations reflect the dynamic nature of an ongoing biology course, ethical issues, technical difficulties, lack of random assignment, and data collection instruments.

These limitations are typically inherent in real classroom research settings and reflect trade-offs for increased ecological validity.

First of all, as a naturalistic setting, the biology classes involved several dynamic interactions among peers, TAs, and the instructor. Although the same instructor taught both sections, each class section had different TAs. Also, despite the same instructor, coherence between the goal of scaffolds and instructional implementation cannot be taken for granted.

In addition, the research design was subject to ethical and practical issues in the field. Ideally, examining the effects of information evaluation would necessitate comparisons between training versus non-training groups. However, in the current study, this control would likely limit students' learning opportunities; consequently, the use of true control groups was not an option in a real classroom setting. Also, due to workload limitations for students and the instructor, task assignments were differentiated as group and individual tasks; thus the study applied different units of analysis to examine the scaffolds effects on argumentation. Group performance provides only a broad estimate of individual performance. Similarly, individual performance reflect only a broad estimate of directs scaffolds effects.

Further, the researcher utilized an open access web-based tool (*Showing Evidence*) as a scaffold. Because the tool was not controllable in the local context, some conflicts with the original research plan were unavoidable. For instance, there were technical errors during student access. Also, changing some functions to fit the study was not feasible for the researcher. For instance, integrating rubrics into the rating system was not possible, nor was adding collaborative scripts to the system. This lack of compliance with research intentions might have lowered the expected effects of the treatments.

The lack of random assignment likely also influenced findings. Due to pre-existing differences among the conditions, the main and interaction effects of the treatments might not have reflected genuine treatment effects. The researcher adjusted possible differences using statistical techniques (i.e., MANCOVA, multilevel analyses).

Finally, the measures and surveys were self-reports that were also administered online, where students could possibly respond to those measures randomly. In order to promote deliberate, reliable responses, however, each measure was delivered as part of classroom activities and assigned grades.

References

Belland, B. R., Glazewski, K., & Richardson, J. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research & Development*, *56*(4), 401–422. doi:10.1007/s11423-007-9074-1

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*(2/3), 191–213.

Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010). Highlighting disputed claims on the Web. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 341–350). New York, NY, USA: ACM. doi:10.1145/1772690.1772726

Iding, M. K., Crosby, M. E., Auernheimer, B., & Barbara, E. (2008). Web site credibility: Why do people believe what they believe? *Instructional Science*, *37*, 43–63. doi:10.1007/s11251-008-9080-7

Kim, B., & Reeves, T. (2007). Reframing research on learning with technology: In search of the meaning of cognitive tools.

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, *75*(3), 285–328.

Lazonder, A. W., & Rouet, J.-F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, *24*(3), 753–765. doi:10.1016/j.chb.2007.01.025

Lee, Y.-J. (2004). Concept mapping your Web searches: a design rationale and Web-enabled application. *Journal of Computer Assisted Learning*, *20*(2), 103–113. doi:10.1111/j.1365-2729.2004.00070.x

Mason, L., Boldrin, A., & Ariasi, N. (2009). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607–633. doi:10.1007/s11251-008-9089-y

Miyata, Y., & Norman, D. A. (1986). The control of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design : new perspectives on human-computer interaction* (pp. 265–284). Hillsdale, N.J.: L. Erlbaum Associates.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. doi:10.1207/s15326985ep4004_5

Rieh, S. Y., & Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 49–71). Cambridge, Mass. : MIT Press, c2008.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, *15*(1), 27. doi:10.1080/10494820600996972

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, *13*(3), 305–335.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Wolfe, J. (2008). Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 141–164. doi:10.1007/s11412-008-9040-x

Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, *58*(1), 181–196.

Zimmerman, B. J., & Tsikalas, K. E. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning? *Educational Psychologist*, *40*(4), 267–271. doi:10.1207/s15326985ep4004_8

APPENDIX A

DEMOGRAPHICS SURVEY

1. What is your gender?
    a. Male
    b. Female
    c. Prefer not to indicate

2. What is your cumulative college GPA? (Enter just two digits, for example: 4.0. If it is your first semester in college, estimate your high school GPA on a 4.0 scale.)

3. What is your major or intended major?
    a. Social Sciences (psychology, anthropology, speech communication, geography, etc)
    b. Humanities (English, Foreign Language, Art, Music, etc)
    c. Sciences (physics, biology, chemistry, geology, pre-nursing)
    d. Mathematics (statistics, mathematics, math education)
    e. Business
    f. Journalism
    g. Education
    h. Agriculture

4. What is your intended minor or area of specialization?
    a. Social Sciences (psychology, anthropology, speech communication, geography, etc)
    b. Humanities (English, Foreign Language, Art, Music, etc)
    c. Sciences (physics, biology, chemistry, geology, pre-nursing)
    d. Mathematics (statistics, mathematics, math education)
    e. Business
    f. Journalism
    g. Education
    h. Agriculture
    i. not applicable

5. How many college level science courses have you already taken?
    a. None
    b. One
    c. Two
    d. Three
    e. Four
    f. Five

6. Are you anticipating taking any more of the following science classes?
    a. No, I am not planning on taking any more science courses (0)
    b. Introductory Biology (BIOL1104)
    c. Anatomy and Physiology
    d. Introductory Geology
    e. Introductory Physics/Astonomy
    f. Introductory Chemistry
    g. Plant Biology or Horticulture
    h. Introductory Ecology
    i. Introductory Entomology

7. How many semesters of college have you *completed* so far?
    a. This is my first semester
    b. 1-2 semesters
    c. 3-4 semesters
    d. 5-6 semesters
    e. 7-8 semesters
    f. More than 8 semesters

8. How many college level math classes have you taken? (Include those you are taking this semester. Do not include remedial math courses.)
    a. None
    b. 1, MATH1113
    c. 1, STAT2000
    d. 1, MATH2200
    e. 2 or more MATH courses
    f. 2 or more STAT courses
    g. 3 or more MATH courses
    h. 3 or more STAT courses
    i. 1 MATH and 1 STAT course

9. Have you taken a research methods course?
    a. yes
    b. no

10. What type of class is most conducive to your learning?
    a. lecture only
    b. activities only
    c. mixture of lecture and activities

11. What type of class do you enjoy most?
    a. lecture only
    b. activities only
    c. mixture of lecture and activities

12. How many years have you used the Internet for information searching?

    a. None
    b. 1-4
    c. 5-9
    d. 10-14
    e. 15 or more

13. During the last month, how often did you search the Internet to find information?
    a. Every few weeks
    b. 1-5 days a week
    c. Once a day
    d. Several times a day

14. How often did you actually find the information you were looking for?
    a. Never
    b. Rarely
    c. Sometimes
    d. Very often
    e. Always

15. How confident do you feel about your searching abilities to find information online for school assignments?
    a. Not at all
    b. Little
    c. Somewhat
    d. Much
    e. Very much

APPENDIX B

PRIOR BIOLOGY KNOWLEDGE TEST

[Macromolecules 3 items]

1. Use this cookie label to answer the next 3 questions:

Nutrition Facts

Serving Size 1 package of cookies (57g)
Servings Per Container 1
(Amount Per Serving)
(Calories 300)
(% Daily Value*)

(Total Fat 16g	25%
Saturated Fat 4g	20%
Trans Fat 4.5g	
Cholesterol 0mg	0%
Sodium 180mg	8%
Total Carbohydrate 37g	12%
Dietary Fiber 1g	4%
Sugars 18g	
Protein 3g)	

(Calories per gram Fat 9 · Carbohydrates 4 · Protein 4)
(INGREDIENTS: Enriched flour, sugar, partially hydrogenated soybean oil, chocolate, cocoa butter, eggs)

1) Which of the following ingredients contains mostly complex carbohydrate?
A) enriched flour B) sugar
C) eggs D) partially hydrogenated sobean oil

2) The Keebler cookie people are attempting to make a healthier version of this cookie by removing the 4.5g of trans fats and replacing it with the same weight in soy protein. How many calories will they save by doing this?
A) 4.5 B) 18 C) 22.5 D) 40.5

3) What ingredient would the Keebler cookie people have to remove to get rid of the trans fats?
A) enriched flour B) sugar
C) cocoa butter D) partially hydrogenated soybean oil

[Cells/Immune System 3 items]

4. Which of the following represents a major difference between prokaryotic cells and eukaryotic cells?
   a. Prokaryotes, not eukaryotes, have cell walls.
   b. Eukaryotic cells tend to have much more extensive inner membrane systems and larger numbers of intracellular organelles than do prokaryotes.
   c. Prokaryotes are unable to carry out aerobic respiration, a process that requires a complex inner-membrane system.
   d. Prokaryotes are a more homogeneous group of organisms than are eukaryotes, which include protozoa, plants, and animals.

5. Vaccinations usually involve injecting ________ into a person.
   a. antibodies against similar diseases
   b. weakened or killed microbes
   c. fully potent disease organisms
   d. antibodies against the disease
   e. antibiotics

6. Which of the following best describes what an antigen is?
   a. a molecule, such as a polypeptide or carbohydrate, that initiates a response from the immune system
   b. a protein in the immune system that is used to identify potential pathogens
   c. a molecule that is used during an immune response to enhance the binding of antibodies to damaged cells
   d. all of the above

[DNA/Cancer 3 items]

7. In figure above, what is being assessed at the end of the first phase of cell growth and differentiation?
   a. complete formation of the spindle
   b. completion of DNA replication
   c. chromosome alignment at the cell equator
   d. completion of cytokinesis
   e. repair of DNA damage

8. The human genome codes for 50,000 to 100,000 proteins, but any given cell might produce only 5,000 to 20,000 different proteins. How is this possible?

   a. Every cell contains a different fraction of the genome.
   b. All possible proteins are made in all cells, but those that are not needed are degraded.
   c. Not every cell has the machinery for transcription and translation.
   d. Genes are regulated so that not all genes are transcribed in all cells.
   e. Some cells use introns to produce proteins, some cells use exons.

9. Which of the following shows the correct matching between the labels (a-e) in the figure above and these terms:

I. nucleotides
II. base pairs
III. chromosome
IV. gene that is used in this cell
V. gene that is not used in this cell

A) I-(b); II-(c); III-(a); IV-(e); V-(d)   B) I-(c); II-(b); III-(a); IV-(d); V-(e)
C) I-(c); II-(b); III-(d); IV-(e); V-(a)   D) I-(b); II-(d); III-(a); IV-(c); V-(e)

[Genetics 3 items]

10. A form of vitamin D rickets, known as hypophosphatemia, is inherited as an X-linked dominant trait if a male with hypophosphatemia marries a normal female, which of the following predictions concerning their potential progeny should be true?

a. All of their sons would inherit the disease
b. All of their daughters would inherit the disease
c. About 50% of their sons would inherit the disease
d. About 50% of their daughters would inherit the disease
e. None of their daughters would inherit the disease

* Use this information for the next 2 questions:

Early onset Alzheimers (A) is caused by a rare autosomal dominant allele that accounts for 3% of all cases of Alzheimers. A woman's mother and maternal grandmother both developed early onset Alzheimers, but none of her other relatives, including her father ever developed the disease.

11. What can you conclude about the woman's mother and father?
a. They are both homozygous for normal allele of the gene.
b. They are both homozygous for the dominant Alzheimers allele.
c. One is homozygous dominant for the Alzheimer's allele, the other is heterozygous.
d. One is homozygous for the normal allele of the gene, the other is heterozygous.
e. They are both heterozygous for the Alzheimers allele.

12. What is the chance that the woman will develop early onset Alzheimers?
A) 0%   B) 25%   C) 50%   D) 75%   E) 100%

[Human Reproduction 3 items)

13. Use this list of important events in the female menstrual cycle for the following question.

1. LH and FSH stimulate follicles so the primary oocyte resumes meiosis I.
2. plummeting progesterone levels signal the disintegration of the endometrium.
3. Follicle cells rupture and release a secondary oocyte.
4. Follicle cells reorganize to form the corpus lutuem.
5. Corpus luteum disintegrates after 7-10 days, halting progesterone production.

What is the correct order of these steps?
A) 5, 3, 2, 4, 1 B) 1, 2, 3, 4, 5 C) 3, 5, 2, 1, 4 D) 2, 4, 5, 1, 3 E) 1, 3, 4, 5, 2

14. Hormones stimulate the uterine lining to thicken in preparation for

a. ovulation
b. fertilization
c. lactation
d. menstruation
e. implantation

15. In human females, fertilization normally occurs in the

a. ovary
c. uterus
e. fallopian tube
g. cervix
i. vagina

[Populations and Invasive Species 3 items)

Galapagos Finches:

Scientists have long believed that the 14 species of finches on the Galapagos Islands evolved from a single species of finch that migrated to the islands one to five million years ago (Lack, 1940). Recent DNA analyses support the conclusion that all the Galapagos finches evolved from the Warbler finch (Grant, Grant & Petren, 2001; Petren, Grant, & Grant, 1999). Different species live on different islands. For example, the medium ground finch and the cactus finch live on one island. The large cactus finch occupies another island. One of the major changes in the finches is in their beak sizes and shapes.

16. What would happen if a breeding pair of finches was placed on an island under ideal

conditions with no predators and unlimited food so that all individuals survived? Given enough time:

a. The finch population would stay small because birds only have enough babies to replace themselves.
b. The finch population would double and then stay relatively stable.
c. The finch populations would increase dramatically.
d. The finch population would grow slowly and then level off.

17. Finches on the Galapagos Islands require food to eat and water to drink.
    a. When food and water are scarce, some birds may be unable to obtain what they need to survive.
    b. When food and water are limited, the finches will find other food sources, so there is always enough.
    c. When food and water are scarce, the finches all eat and drink less so that all birds survive.
    d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.

18. Once a population of finches has lived on a particular island for many years,
    a. the population continues to grow rapidly
    b. the population remains relatively stable, with some fluctuations.
    c. the population dramatically increases and decreases each year.
    d. the population will decrease steadily.

APPENDIX C

LIST OF INQUIRY TASKS

Topics			Description
Group Task	Common Task	Burning Controversy	Students investigate claims about which molecules build up in muscle cells causing them to burn during exercise
	Chosen Task	Vaccination Debate	Students create media reports to convince the public either to use or avoid one vaccine that is currently available. Students investigate claims about which molecules build up in muscle cells causing them to burn during exercise
		Worst Food in America	Students create a media story about the worst and best example of a food item found at a grocery store or restaurant.
		Cancer: What are my options	Students assume the role of a loving family member investigating the mode of action and risk benefits of a specific chemotherapy drug.
		Genetic Testing	Students assume the role of helping a loved one make an informed decision about the likelihood of having inherited gene that may cause a disease.
		Reproduction Myth Buster	Students use their knowledge of human reproduction to create a media piece to correct a major misconception about human reproduction.
		Fish Futures	Students create flyers to post at a campus eatery to inform consumers of whether they are purchasing fish from sustainable stocks.
Individual Task	Pre-Task	To Use or Not to Use Plastic Ware	Students investigate and write a short essay that answers whether using plastic ware is harmful.
	Post-Task	Cholesterol Control	Students assume the role of helping a family member diagnosed with a higher cholesterol level make an informed decision whether to consent to taking prescription medication.

APPENDIX D

INFORMATION EVALUATION TRAINING MATERIAL

Expert's Rank	Your Rank	Web Site
3		Food Additives ~ CSPI's Food Safety Shopping was easy when most **food** came from farms. Now, factory-made **foods** have made chemical **additives** a significant part of our diet. In general, it's best to ...www.cspinet.org/reports/chemcuisine.htm
2		Food Additives - Food and Drug Administration Links to Information about **Food Additives** and **Food Additive** Petitions. ...www.fda.gov/food/foodingredientspackaging/foodadditives/default.htm
5		Food Additive Essentials **Food Additive** Essentials. The meals we are consuming nowadays are very far from being natural. And, of course, we are not getting all the necessary vitamins ... ...www.foodadditivetips.com/
1		Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial BACKGROUND: We undertook a randomised, double-blinded, placebo-controlled, crossover trial to test whether intake of artificial food colour and **additives** (AFCA) affected childhood behaviour. METHODS: 153 3-year-old and 144 8/9-year-old children were ... D. McCann, A Barrett, A Cooper, D Crumpler… - The Lancet, 2007 - Elsevier
4		Understanding Food Additives An educational site to support teaching of **Food Additives**, E-Numbers and related topics. www.understandingfoodadditives.org/
6		food reactions.org: Food Additives **Food additives** have been used for centuries; for example, preserving food by pickling (with vinegar), salting, as with bacon and dried tomatoes, or using sulphur ... www.foodreactions.org/allergy/additives/

Note. Students across conditions were trained to rate six given websites about Food Additives using CRAP testing. In addion, the annotation tool users were provided an additional annotation tool use training material (Appendix E).

APPENDIX E

INTEL *SHOWING EVIDENCE* TOOL TRAINING MATERIAL

1. Screen Shot

2. Steps to Follow

Steps	Description
1. Create claim	Create a claim and describe your explanation. [Example] • Claim: Radon gas is seeping into the house, colliding with other air molecules. The family is getting cancer from breathing in the radon. • Explanation: There are several factors that support radon sources: bedrock, stone facade of the home, bricks. Being a new house, it is air tight.
2. Create evidence	Record information in the evidence bin: put a summary title, describe explanation, and record the source. [Example1] • Summary: Radon comes from rock/soil • Explanation: Radon gas comes from radioactive decay of radium, a ubiquitous element found in rock and soil. It moves from soil into the air, emits alpha, beta particles, and gamma rays. Radiation damages cells & results in cellular transformation in the respiratory tract, which can lead to radon-induced lung diseases or cancer. • Source: http://www.atsdr.cdc.gov/HEC/CSEM/radon/ [Example2] • Summary: Average level of radon is 4 picocuries • Explanation: No federal or state standards define the amount of radon that is safe. The EPA has set a level of concern or "action level" for homes, above which remedial action should be considered. This radiation level is currently an annual average of 4 picocuries per liter of air (pCi/L). • Source:http://www.oag.state.ny.us/environment/radon96.html#safe_level
3. Rate evidence quality	Evaluate the source of information using a 5-point rating system. Important questions are "Do I have credible and accurate source of information?" [Example1] • Quality Rating: ✓✓✓✓ • Rating Rationale: Agency for Toxic Substances & Disease Registry appears reputable. They work with the EPA and US Health & Human Services. [Example2] • Quality Rating: ✓✓✓✓ • Rating Rationale: The New York Attorney General submitted this information. He is obviously a well-respected individual.

Steps	Description (Continued)
4. Rate evidence strength	Judge whether evidence identified from the sources supports (+) or weakens (-) the claim and whether the evidence is relevant, important, and valid. Then put your reasoning. It serves as a warrant statement. Important questions are: (1) "Is the evidence relevant to the claim? (How does the evidence relate to the claim?") (2) "Is the evidence central to the claim?" (3) "How did the author generate this evidence? (Research methodologies)" "Is the evidence generalizable? Are there any variables unexplained?" [Example1] • Support Rating: + + + • Reasoning: Sally's house facade is covered with rock from a local quarry. Her house is built on a hilltop. Radon could be seeping from the rocks & infiltrating the house with radon gas. The evidence moderately supports the claim. [Example2] • Support Rating: − − − − • Reasoning: This evidence shows that the level of radon in Sally's house is only average.
5. Rate the strength of claim	Evaluate the entire information collection. Based on the summative evaluation, rate the strength of your claim. Important questions are: (1) "Do I have qualified and strong evidence?" (2) " Do I have converging evidence across different sources" (3) "Are the important counter-arguments explored? Can you explain why they are wrong?" [Example] • Support Rating: ** • Reasoning: Radon does not seem a likely source for Sally's illness. The primary symptom she has in common with radon is her hair is falling out. But, that only occurs with radiation treatment of cancer patients, & Sally has not had radiation treatment.
6. Draw a final conclusion	Create another claim or seek for another source of information. Based on several reasoning, you may draw a final conclusion.

Note. The screen shot and examples are taken from Intel (https://educate.intel.com/ workspace/tryit/

SEtryit2.aspx?LID=en)

APPENDIX F

CHECKLIST

Item	Questions
Currency	1. Can you tell how recently the site has been updated?
Reliability	2. Is the information based on scientific evidence? • Is evidence provided or reported for claims? • Are scientific peer reviewed journals cited? • Is this information likely to be evaluated well by informed scientists?
	3. Is there similar information given across reliable sources? • Do you have multiple sites or authors that give the same information? • Is information in a site contradicts other sites that you think are trustworthy? • Is the account complete? Or does it omit information that other reliable sources mention?
	4. How well does the site explain the information? • Do you understand how the process works based on the information provided? • Does the explanation fit together with your prior scientific knowledge or with information from other reliable sites? • Considering other trustworthy sites, does each interpretation of the evidence fits together to generate a coherent explanation of a scientific phenomena?
Authority	5. Does the person who is providing the information have the credentials and knowledge to provide reliable information?
Purpose	6. What is the motivation for providing this information? • Is the site pushing a specific agenda? Are they trying to sell you something or elicit a donation? Do they have a political agenda?

Note. The CRAP test was revised reflecting preliminary study findings (Meola, 2004; Metzger, 2007; Wiley et al., 2009) since the original CRAP focus primarily on surface features related to source quality not addressing how source information is generated, justified, and related to claims and other pieces of information (c.f., Mason et al., 2009).

APPENDIX G

PERMISSION TO USE INTEL'S *SHOWING EVIDENCE* ANNOTATION TOOL

From: teacher.training@intel.com
Subject: Re: Subject: Request for Permission to Use Showing Evidence Tool
Date: October 25, 2011 6:28 PM EST
To: cotton93@uga.edu

Dear Ms. Kim,

Thanks for getting back to us.

We forwarded your request to the Intel Education staff for review and they said your intentions for the showing evidence tool are within normal usage expectations. We think it's great that you will be leveraging the thinking tools for your research methods. The staff would greatly appreciate knowing the results of your dissertation when it's completed if you don't mind sharing!

We wish you the best with your research and completion of your doctorate. If you have any questions or concerns, please contact us at teacher.training@intel.com.

Thanks again!

Casey Rood

----- Original Message -----

Dear Intel® Education Representative:

My name is So Mi Kim, and I am a Ph.D. student in Learning, Design, and Technology program at The University of Georgia. I am currently working on my dissertation study, which aims to enhance college students' science literacy skills by incorporating real-world Biology problems as part of a course to fulfill a general education requirement. I would like to use showing evidence tool and rubrics—quality and strength of evidence rubrics and project assessment rubrics—located at http://educate.intel.com/en/ThinkingTools/ShowingEvidence/, which I will use to scaffold college students in analyzing and evaluating information.

Training in information literacy has been assumed to improve critical thinking and content understanding. However, its systematic investigation has been slow. Responding to such needs, my study will take place in one undergraduate general biology class of sample size 300. I have another control group of the same size. The study specifically focuses on 1) information evaluation and integration aspects of information literacy during scientific argumentation and 2) their scaffolding with a focus on metacognition. Research hypothesis is that metacognitive scaffolding positively influences information evaluation and integration, which again positively influences students' argumentation skills and deep understanding. Showing evidence tool will be utilized to visualize student thinking process and thus activate student metacognition.

I am a certified master teacher trained at Intel®Teach Thinking with Technology Master Teacher Course, which was pilot-tested in Austrailia in 2006. In addition, I had been involved in localizing that course in Korea that I am quite experienced in using Intel thinking tools and familiar with philosophies and principles underlying the tools. Base on these experiences and understanding, I would like to contribute to improving students' functional science literacy at a college level. I believe that my findings will also benefit your educational initiatives over the world by sharing evidence of the program effectiveness at a college level.

Findings will be presented at major conferences and published in major journals in the area of science education and instructional technology. In any case, I will cite Intel as the source and include any other information that you would like me to include to properly crediting the creator. I respect copyrights that Intel reserves and will follow the regulations. If you have any further questions, feel free to talk to me. I appreciate your consideration in advance.

So Mi Kim

Ph.D. Student, Learning, Design, and Technology, University of Georgia, 614 Aderholt Hall, Athens, GA 30602 USA, TEL: +1 706 224 1038
Assistant Editor, Educational Technology Research & Development
cotton93@gmail.com, cotton93@uga.edu

APPENDIX H

GOAL INSTRUCTIONS

1. An Excerpt from Group Task Instruction Form (A) for a Persuasion Goal

Group Task 1: Burning Controversy

Assignment: You know what your trainer told you about why your muscles burn while exercising, but how do they know what is causing this phenomenon? You assignment will be to investigate this claim and compose a scientific argument supporting your understanding.

- **Research claims**: Which molecules build up in muscle cells causing them to burn? Are these compounds good or bad? Do a quick Google search and come up with your claim. Use valid sources that you find to help you come up with your own opinion about this issue. **<u>Assume that you convince your friends of your claim</u>**.

2. An Excerpt from Group Task Instruction Form (B) for a Balanced Reasoning Goal

Group Task 1: Burning Controversy

Assignment: You know what your trainer told you about why your muscles burn while exercising, but how do they know what is causing this phenomenon? You assignment will be to investigate this claim and compose a scientific argument supporting your understanding.

- **Research claims:** Which molecules build up in muscle cells causing them to burn? Are these compounds good or bad? Do a quick Google search to come up with what people might be saying about this. **<u>You may find several sides to the story. Use valid sources that you find to help you come up with your own opinion about this issue. Make sure that you critically examine different sides and construct your own claim on the issue</u>**. Assume that you will report as a health column editorial of a newspaper.

3. An Excerpt from Individual Task Instruction From (C) for a Neutral Goal Statement

Individual Task 1: To Use or Not to Use Plastic Ware

Assignment: You will investigate and write a short essay (no fewer than 300 words except references) that answers the following question "Is using plastic ware (e.g., drinking out of plastic bottles or microwaving food with plastic wrap) harmful?"

Write your opinion based on your understanding. <u>Assume that you would inform your peers of this topic.</u> Can you use the evidence you have found to make recommendations? Your essay will be graded based on the quality and thoroughness of argumentation.

APPENDIX I

INFORMATION EVALATION BEHAVIOR SCALE

[Scale]

False										True	Not Applicable
0	10	20	30	40	50	60	70	80	90	100	☑

Expectation set about needed sources (Predictive Evaluation; 3 Items)

1. I was eager to find a single, right website that contained useful information.
2. I expected to refine my search goals after a series of searches.
3. I asked myself whether different perspectives might exist in relation to the given topic.

Evaluation of innate merit of sources (Intrinsic Evaluation; 5 Items)

1. I checked whether information source was outdated.
2. I asked whether the author had proper authorities.
3. I checked whether the source had proper support.
4. I differentiated primary, secondary, and tertiary sources of information.
5. I asked whether the source had any hidden propagandas (e.g., sales promotion).

Evaluation of relative strength and utility of sources (Extrinsic Evaluation; 5 Items)

1. I asked whether the source was relevant to my claim.
2. I asked how strongly the source supported my claim (e.g., research methods and details).
3. I checked whether the source claim agreed to other sources.
4. I intentionally brought up counter-evidence to check the validity of my claim.
5. I weighed different sources to strengthen my claim.

Summative evaluation of source collections (Reflective Evaluation; 3 Items)

1. I periodically paused to draw an overall picture of multiple sources I had selected
2. I re-evaluated my assumptions if I found the sources did not fit my understanding.
3. I tried to translate multiples sources into a new whole.

* Items were randomly presented to each student.

APPENDIX J

PERCEIVED TASK VALUE: SCIENCE MOTIVATION SUB-SCALES

[Scale]

Never	Rarely	Sometimes	Usually	Always
1	2	3	4	5

Intrinsically motivated science learning items (Interest; 5 Items)

1. I enjoy learning science.
2. The science I learn is more important to me than the grade I receive.
3. Learning science is interesting.
4. I like science that challenges me .
5. Understanding the science gives me a sense of accomplishment.

Extrinsically motivated science learning items (Importance; 5 Items)

1. I like to do better than the other students on the science tests.
2. Earning a good science grade is important to me.
3. I think about how learning science can help me get a good job.
4. I think about how my science grade will affect my overall grade point average.
5. I think about how learning the science can help my career.

Relevance of learning science to personal goals (Utility; 5 Items)

1. The science I learn relates to my personal goals.
2. I think about how the science I learn will be helpful to me.
3. I think about how I will use the science I learn.
4. The science I learn is relevant to my life.
5. The science I learn has practical value for me.

* Items were randomly presented to each student.

APPENDIX K

WORKING MEMORY CAPACITY TEST

[Introduction] Your job in this test is to memorize the words you see on the screen while you also solve math problems. You can take as long as you want to answer the math problems correctly, but DO NOT write down the words you see. Here is an example of what the math problems are going to look like.

Is (2 x 1) + 1 = 2?

As soon as you see one of these problems appear on the screen, I’d like you to read the equation, then verify if the answer is correct or not by clicking YES or NO. When you have clicked ">>" you will be presented with a word:

dog

Say the word out loud and then wait to be continued to the next math question. At the end of a set, you will see a question like this:

Type in all of the words that you saw in the set

When you see this question, your job is to type in all of the words that you saw in that set. I’d like you to type them in the same order that you saw them in, one word on each line. If you can't remember all of the words, leave the space for the word(s) you can’t remember blank. In this case you would only enter one word, “dog” but later you will have to memorize and enter more than one word.

[Practice] Let’s begin with some practice, so you can get used to how this works. Answer this practice question, then click ">>" Say the word on the next page and wait to be continued. Then answer the next math question and say the second word. Then type in the words when you are asked to do so.

IS (7 x 1) - 3 = 3?........ **cheek**
IS (8 / 2) + 4 = 2?........ chalk
Type in all of the words that you saw in the set

Okay, that was easy wasn't it? Here are some more practice questions before we get into the actual test.
IS (6 x 3) + 2 = 17 plant
IS (3 / 1) - 2 = 3 foot
Type in all of the words that you saw in the set

Okay, do you think you are ready to begin the real test or do you need some more practice?

- **I am ready to begin the real test**
  (when clicked on, students will be provided a main test).
- **I need some more practice first**

Okay, here is some more practice:

IS (8 / 2) - 1 = 3 bike
IS (10 / 10) - 1 = 2 ball
Type in all of the words that you saw in the set

[Main Test] Now it is time to start the working memory test. The real test is going to work just like in practice, but there are going to be a different number of math problems and words in each set. Sometimes there will be 2 just like in practice, but other times there will be 3, 4, 5, or 6 in a set. The order of these sets is random, so you won't know how big a set is until you're finished with it. Just like in practice, if you can't remember a word or words, please leave a space blank for it.

Remember, you may take as long as you need to solve the math problems, DO NOT write down the words.

(click ">>" to begin the test)
IS (10 x 1) - 7 = 3 y clouds
IS (10 / 1) + 1 = 10 n baby
IS (9 x 3) + 2 = 27 n sand
Type in all of the words that you saw in the set

IS (10 x 2) - 1 = 19 y chance
IS (4 / 1) - 3 = 1 y end
IS (5 x 2) + 2 = 12 y course
IS (8 x 1) + 2 = 10 y floor
IS (7 x 1) + 6 = 13 y soil
Type in all of the words that you saw in the set

IS (7 / 7) + 5 = 6 y hair
IS (10 / 2) + 4 = 3 n state
IS (9 / 3) - 2 = 1 y bush
Type in all of the words that you saw in the set

IS (4 / 1) + 1 = 4 n mind
IS (7 x 2) - 1 = 14 n fact
Type in all of the words that you saw in the set

IS (2 x 3) + 1 = 4 n cot
IS (4 / 2) + 1 = 6 n mold
IS (6 / 2) - 1 = 1 n class

IS (9 / 1) + 8 = 18 n hill
IS (6 / 2) - 2 = 2 n jar
Type in all of the words that you saw in the set

IS (8 x 4) + 2 = 34 y form
IS (6 x 2) - 2 = 10 y east
IS (7 x 7) + 1 = 49 n ground
IS (8 / 4) + 6 = 8 y check
IS (3 / 1) + 3 = 6 y bench
Type in all of the words that you saw in the set

IS (7 / 1) + 2 = 7 n map
IS (6 / 6) + 2 = 4 n pipe
IS (10 x 1) - 5 = 10 n side
IS (5 x 1) - 1 = 4 y heart
IS (2 x 1) - 1 = 1 y ears
IS (9 / 3) + 3 = 6 y world
Type in all of the words that you saw in the set

IS (10 / 1) + 3 = 13 y face
IS (10 x 2) + 2 = 21 n jail
IS (9 x 3) - 2 = 25 y point
IS (2 / 1) - 1 = 1 y lamp
Type in all of the words that you saw in the set

IS (9 x 1) - 5 = 5 n drill
IS (10 / 2) + 4 = 9 y flute
IS (3 x 2) + 1 = 6 n rain
IS (5 / 5) + 4 = 5 y town
IS (2 x 4) + 1 = 8 n sea
IS (8 / 8) + 1 = 2 y hat
Type in all of the words that you saw in the set

IS (10 / 5) - 1 = 1 y beach
IS (10 / 1) - 5 = 4 n rat
Type in all of the words that you saw in the set

IS (4 x 4) + 1 = 17 y lot
IS (9 / 1) + 4 = 14 n cone
IS (6 / 2) - 2 = 2 n kid
IS (9 x 1) + 9 = 1 n tin
Type in all of the words that you saw in the set

IS (8 x 1) + 5 = 13 y grass
IS (6 x 2) - 3 = 10 n oil
IS (8 / 4) - 1 = 1 y ice

Type in all of the words that you saw in the set

IS (3 / 3) + 5 = 14 n bear
IS (10 / 1) + 2 = 12 y box
Type in all of the words that you saw in the set

IS (10 / 1) - 1 = 9 y church
IS (6 x 1) - 4 = 1 n table
IS (10 / 2) - 3 = 2 y jam
IS (5 x 2) - 5 = 4 n move
IS (2 x 1) - 3 = 6 n iron
IS (7 / 1) - 2 = 7 n branch
Type in all of the words that you saw in the set

IS (3 x 1) + 2 = 6 n half
IS (7 / 1) + 6 = 12 n ants
IS (6 / 3) + 9 = 11 y gold
IS (5 x 1) - 1 = 5 n hole
Type in all of the words that you saw in the set

APPENDIX L

ARGUMENTATION QUALITY SCORING RUBRICS

1. Evidence Quality (EQ)

Level	Criteria	Description
0	No Evidence	Students provide no evidence or topically irrelevant evidence.
1	Unqualified Evidence	Students provide unqualified evidence, which is outdated, unreferenced (e.g., news without source information), unauthorized, or biased.
2	Qualified Evidence	Students provide qualified evidence, which is, however, single or piecemeal evidence.
3	Converging Evidence	Students provide multiple, converging sources of qualified evidence that agree to support a claim.
4	Alternative Evidence	Students consider alternative arguments, but do not articulate why alternative claims/evidence are wrong.
5	Dialectic Evidence	Students provide relevant, qualified, and converging evidence from multiple sources. Students also provide counter-evidence and its rebuttal.

Note. Evidence quality was created reflecting a rhetorical structure of argumentation (Clark & Sampson, 2008; Kuhn, 1989; Toulmin, 1958). Scores ranged from 0 to 5.

2. Reasoning (R)

Level	Criteria	Description
0	Unjustified	Students provide no explanation.
1	Unsubstantiated	Students provide an explanation, but evidence is not incorporated.
2	Substantiated	Students provide an explanation articulating sources of evidence. However, it is not clear how strongly their source of evidence supports a claim or how sources relate to one another.
3-4	Warranted	Students provide warrant statements explicitly commenting on evidence quality (qualifiers) or/and inter-relations among evidence sources (linkages)

Note. Reasoning scoring scheme was created based on grounded reasoning development framework (Brem & Rips, 2000; Kuhn, 1989). Scores ranged from 0 to 4.

3. Conceptual Integration for Group Tasks (CIG)

Level	Criteria	Description
0	Pre-structural	Misconceptions dominate student argumentation.
1	Uni-structural	Students can deal with one single aspect correctly showing limited understanding.
2	Multi-structural	Student response focuses on several relevant aspects, but they are treated independently and additively. Some terms may not be concisely or clearly defined. Too little or irrelevant detail is included to explain relations.
3	Relational	Different aspects have become integrated into a coherent whole. Sufficient detail is provided to explain relations in a correct context.
4	Extended Abstract	Student response has been conceptualized at a higher level of abstraction. Students can theorize or/ evaluate alternative views.

Note. Conceptual integration scoring scheme for group tasks was revised from Biggs and Collis (1982). Scores ranged from 0 to 4.

4. Conceptual Integration for Individual Tasks (CII)

Criteria		0	1	2
1. Substance (e.g., plastics)		Students do not differentiate different kinds.	Students specify one kind.	Students specify two or more kinds.
2. Process	2-1. Cause (e.g., heat)	Students do not specify agents/conditions of events.	Students specify agents/conditions of events in a correct context	-
	2-2. Event (e.g., Estrogen Mimics)	Students do not detail what happens.	Students provide partial details of events.	Students provide sufficient details of events.

Note. Conceptual integration scoring scheme for individual tasks is analytic to sum component scores. Criteria were created based on science concept categories(Chi et al., 1994). Summed scores ranged from 0 to 5.

APPENDIX M

STUDENT TASK PERCEPTION ATTRIBUTES LIST

Category	Sub-Category	Definition	Excerpts
Goal	[g1] Fixed Science	- Define goals as finding facts/expected answers.	"…one certain answer we had to find"
	[g2] Open Science	- Define goals as arguments and testing theories	"I understood the task to be a debate"
Task Focus	[t1] Learning Task	- Define main tasks as learning specific contents	"It was learning about…"
	[t2] Research Task	- Define main tasks as searching for, and organizing information to support claims	"…do research, find, and compose it all into a draft"
	[t3] Evaluation Task	- Define main tasks as backing ideas with reliable sources	"My concern was to find reliable sources to back our claim"
	[t4] For-Grades Task	- Define main tasks as meeting instructor expectations or following instructions	"We needed to find what (the instructor) think is the cause of …"
Evaluation Strategies	[e1] Easy Evaluation	- Use known, given, or easy-to-access sources	"I decided to stick to 1-2 major sources…"
	[e2] Intrinsic Evaluation	- Use CRAP strategies or focus on primary sources	"My concern was finding scholarly articles"
	[e3] Extrinsic Evaluation	- Compare and contrast sources	"We need to compare findings and ideas.."
	[e4] Reflective Evaluation	- Change strategies or views while addressing complicate topics	"I was originally an easy Googler, but …enlightened to find (this task) took more digging through the murky information"
Perceived Task Value	[v1] Cost	- Complain required time and efforts	"It was just a busy work"
	[v2] Ease	- Consider the task doable	"It was pretty much straightforward"
	[v3] Difficulty	- Consider the task challenging	"We had trouble deciding what were supposed to do…"
	[v4] Topical Interest	- Relate the task to personal interests and knowledge	"I liked to learn about…"
	[v5] No Interest	- Consider the task demotivating	"I was not motivated to do…"
	[v6] Utility	- Consider the task useful	"We can apply this to everyday lives"
	[v7] No Utility	- Consider the task unnecessary for college education	"It was pointless…we already learned it from high school research"
Background knowledge	[p1] Prior Knowledge	- Have ideas, heard, or learned about the given topic	"I had basic understanding about…"
	[p2] No Prior Knowledge	- Have no related knowledge	"We had learned very little about this""]

APPENDIX N

TASK PERCEPTION CLUSTERING RESULTS AND CLUSTER PROFILES

Goals	Cluster	%	Cluster Attributes (%)																		
			g1	g2	t1	t2	t3	t4	e1	e2	e3	e4	v1	v2	v3	v4	v5	v6	v7	p1	p2
Persuasion	1	23	7.5	17.5	**97.4**	15.3	15.5	7.7	0.0	20.8	14.3	0.0	60.0	0.0	11.4	**80.0**	0.0	20.0	**57.1**	0.0	0.0
	2	32	**45.3**	**0.0**	**2.6**	27.8	**0.0**	**53.8**	**66.7**	4.2	28.6	0.0	20.0	50.0	40.0	10.0	0.0	20.0	42.9	0.0	33.3
	3	23	34.0	**0.0**	**0.0**	33.3	**60.3**	7.7	33.3	**75.0**	28.6	0.0	0.0	14.3	22.9	0.0	0.0	0.0	0.0	0.0	44.4
	4	23	13.2	**82.5**	0.0	23.6	24.1	30.8	0.0	0.0	28.6	**100**	20.0	35.7	25.7	10.0	0.0	60.0	0.0	**100**	22.2
	Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.0	100	100	100	100
Balanced	1	33	29.7	**0.0**	**61.8**	**0.0**	18.5	33.3	0.0	5.3	0.0	6.7	66.7	0.0	**0.0**	46.7	66.7	0.0	50.0	0.0	25.0
Reasoning	2	28	**40.5**	0.0	20.6	**63.2**	32.3	**55.6**	0.0	0	0.0	0.0	0.0	18.8	0.0	0.0	0.0	0.0	50.0	**100**	0.0
	3	12	2.7	0.0	2.9	7.4	7.7	0.0	**100**	10.5	0.0	**46.7**	33.3	**62.5**	**68.0**	0.0	0.0	33.3	0.0	0.0	**75.0**
	4	14	13.5	**88.0**	2.9	10.3	15.4	11.1	0.0	0.0	0.0	**33.3**	0.0	12.5	16.0	6.7	33.3	0.0	0.0	0.0	0.0
	5	13	13.5	12.0	11.8	19.1	**26.2**	0.0	0.0	**84.2**	**100**	13.3	0.0	6.2	16.0	**46.7**	0.0	66.7	0.0	0.0	0.0
	Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Note. * denotes important attributes for the creation of one cluster separate from another ($p<.05$). G1 and G2 indicate persuasion and balanced goal conditions respectively. [g1] Fixed Science, [g2] Open Science, [t1] Learning Task, [t2] Research Task, [t3] Evaluation Task, [t4] For-Grade Task, [e1] Easy Evaluation, [e2] Intrinsic Evaluation, [e3] Extrinsic Evaluation, [e4] Reflective Evaluation, [v1] Cost, [v2] Ease, [v3] Difficulty, [v4] Interest, [v5] No Interest, [v6] Utility, [v7] No Utility, [p1] Prior Knowledge, [p2] No Prior Knowledge