Learning and Instruction 57 (2018) 1–4

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Introduction to the special issue: Desiderata for a theory of multi-source multi-modal comprehension

Jennifer G. Cromley

University of Illinois at Urbana-Champaign, United States

ABSTRACT

In this special issue, contributing authors consider work at the intersection of two relatively established lines of research: multi-source (e.g., multiple texts) comprehension and multi-modal (e.g., narrated video) comprehension. This is a challenging space in which to work, as there is great complexity in the theories and empirical base in each area. Commonalities and important differences across the two literature are highlighted, together with a set of criteria that any emerging theory of multi-source, multi-modal comprehension should account for: cognitive learning strategies and metacognitive monitoring, which may interact with media; affect and other individual differences; and task effects.

1. Introduction

In this special issue, contributing authors consider work at the intersection of two relatively established lines of research: multi-source (e.g., multiple texts) comprehension and multi-modal (e.g., narrated video) comprehension. This is a challenging space in which to work, as there is great complexity in the theories and empirical base in each area—multi-source comprehension and multi-modal comprehension. A general finding from multi-source research is that forming a mental model that integrates information across texts requires substantial effort and the utilization of deep processing strategies (Hagen, Braasch, & Bråten, 2014). A general finding from the multi-modal processing research is that factual learning from each modality is perhaps less demanding for many learners, but combining information across modalities is likely difficult for most (Schnotz, 2014). Across both literature, teaching students specific skills and learning strategies (e.g., multi-source strategies such as sourcing, multi-modal strategies such as linking terms from text to diagram; the Historical Documents Matrix used by Van Meter & Cameron, this issue) is often associated with improvements in mental model construction.

These two areas of research differ somewhat in their attention to processing during learning—this has been a major focus of the multi-source comprehension literature, with its extensive use of think-aloud protocols and identification of multiple-documents-specific strategies such as sourcing and corroboration (Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). The multi-modal comprehension literature has only occasionally focused on processes during learning (e.g., think-alouds, eye tracking, computer log files), but has much more often tested effects on learning from specific modifications to the media presentation (Mayer, 2014). The contributions to this special issue all analyze the process of learning in order to better understand how learners form mental models from these complex stimuli.

2. Desiderata for a theory of multi-source, multi-modal comprehension

I lay out a set of criteria that any emerging theory of multi-source, multi-modal comprehension should account for: cognitive learning strategies and metacognitive monitoring, which may interact with media; affect and other individual differences; and task effects.

2.1. Cognitive and metacognitive strategies

Researchers have documented a range of cognitive (e.g., summarizing) and metacognitive (e.g., feeling of knowing) strategies used across both the multi-source comprehension literature and the multi-modal comprehension literature. Not surprisingly, there is substantial overlap in the strategies found in both literature—and indeed, there is substantial overlap with the text comprehension literature generally (Cerdán & Vidal-Abarca, 2008; Stadtler & Bromme, 2007). One set of strategies uniquely associated with the multi-source comprehension literature is related to sources. *Sourcing* is defined as identifying which source provided a particular piece of information (Wineburg, 1991), and is related to *evaluation of source credibility* (Rouet et al., 1997), which involves judgments of how the specific source may bring a particular perspective to the writing (e.g., democratic versus republican

E-mail address: jcromley@illinois.edu.

https://doi.org/10.1016/j.learninstruc.2018.02.004
Received 9 January 2018; Received in revised form 18 February 2018; Accepted 21 February 2018
Available online 07 March 2018
0959-4752/

sources writing about the Vietnam War; industry versus scientific sources writing about risks from a specific technology). Strategies associated with multi-modal comprehension are focused on visual representations such as *coordinating informational sources* (e.g., Azevedo & Cromley, 2004), *visual search* (e.g., Gegenfurtner, Lehtinen, & Saljo, 2011), and *inferences* within diagrams (e.g., Cromley, Snyder-Hogan, & Luciw-Dubas, 2010). These strategies may vary depending on the specific modality (see List, this issue); for example, video/animations tend to present a very large amount of information in a very short time. Visual search may therefore be both more challenging and more useful in video/animations than in static diagrams. Likewise, user controls such as a rewind button or slider bars (Scheiter & Gerjets, 2007) apply only to certain media types, such as video/animations, serious educational games, and most computer-based simulations. Contradictory results are also found for higher-level strategies (i.e., more knowledge-transforming strategies such as summarizing versus paraphrasing) in text versus in other representations. In some cases, the non-textual representations seem to prompt higher-level strategy use (Cromley et al., 2010) and in other cases the text prompts higher-level strategy use (Van Meter and Cameron, this issue).

2.2. Individual differences

The literature on multi-source comprehension and in multi-modal comprehension has often highlighted the role of individual differences. For example, more sophisticated epistemic beliefs, higher prior knowledge, and more expertise improve multi-source comprehension (Braasch, Bråten, & McCrudden, 2018). Readers with beliefs about knowledge being complex and changing are more likely to notice contradictions between texts from different sources (Kammerer, Kalbfell, & Gerjets, 2016). Readers with more prior knowledge are also likely to compare and contrast differences in historical accounts ("what happened?") and different explanations for why those events took place ("why did it happen?"; Wineburg, 1991). Likewise, expertise in a domain is associated with very different patterns of reading from multi-source sets such as legal experts reading case law versus non-lawyers reading the same cases (Lundeberg, 1987).

In multi-modal learning, higher prior knowledge, spatial skills, and working memory capacity have all been shown to improve comprehension (Mayer, 2014). As with text comprehension, knowing more about a topic presented in a multi-modal format appears to help with forming an elaborated mental model of the topic (Cook, Carter, & Wiebe, 2008). Since many illustrated topics have spatial features that relate to the depicted system (e.g., right and left sides of the heart have different functions and hence different structures; the left ventricle is heavily muscled unlike the right ventricle), it is no surprise that students who score higher on measures of spatial skills also show better comprehension of multi-media, but with complex mappings of specific spatial skills to specific stimuli (Atit, Gagnier, & Shipley, 2015).

Likewise, higher working memory capacity is associated with better comprehension of multi-modal texts, given the high information density of multi-modal displays (Sanchez & Wiley, 2006). This is especially true when complex span tasks are used to measure working memory. Finally, scaffolds designed to help low-knowledge, low-spatial, or low-WM learners can "backfire" (termed the expertise reversal effect) and harm high-knowledge, high-spatial, or high-WM learners (Kalyuga, 2014). Any theory of multi-source, multi-media comprehension should take account of individual differences, but more work will be needed to include the frequently-used individual differences from both multi-source and multi-modal research.

2.3. Tasks assigned to learners

A third factor that a theory of multi-source, multi-modal documents comprehension will need to account for is the task set for the learner. For example, the tasks assigned to learners in the studies reported here range from reading to answer main idea and factual questions (Singer Trakhman, Alexander, & Silverman, this issue), to reading and taking notes using a Historical Document Matrix (Van Meter & Cameron, this issue), to judging how supportive an argument diagram is for a claim (van Amselvoord & Schilperoord, this issue), to annotating text or videos (List, this issue), to judging whether representations match or do not match (Schüler, this issue).

While general reading comprehension research has taken on different purposes for reading, neither the multi-text nor the multi-modal comprehension literature have considered the learning task to any great extent. This is unfortunate given some findings from the general comprehension literature: for example, students asked to read for the purpose of taking a test tend to use more low-level strategies and to recall isolated facts, whereas those asked to read for the purpose of explaining what they read to a peer tend to use more high-level strategies and to recall interconnected information from what was read (McCrudden, Magliano, & Schraw, 2011). Given the wide range of tasks in the multi-text and multi-modal literature, it remains to be seen how the specifics of various learning tasks affect different learning outcomes.

2.4. The challenge of strategy, task, and learner interactions

Although it has been under-researched in multi-source research, the larger multi-modal comprehension research often enough shows interaction effects among these factors. For example, some interventions privilege learners with certain characteristics and disadvantage others. In my own research, I have found that a learner-completed drawing task advantaged high-spatial students and disadvantaged low-spatial students, who under-scored those in a non-drawing condition (Cromley et al., 2013). Likewise, some powerful learning strategies depend on prior knowledge (e.g., elaborative inferences), and therefore advantage learners who score high on this individual difference. For example, cueing in text-and-diagrams is more beneficial for low-knowledge learners than for high-knowledge learners (e.g., Johnson, Ozogul, & Reisslein, 2015). A comprehensive theory of multi-source, multi-modal comprehension will also have to account for interactions among these factors; null results may not in fact reveal lack of an effect in our research, rather, they may hide effects of a missing moderator.

3. Some design challenges for multi-source, multi-modal comprehension research

Choosing sets of stimuli poses a challenge for designing research studies in multi-source, multi-modal comprehension. In the multi-source literature, texts are chosen with contradictory and incompletely overlapping information; in the multi-modal comprehension literature, usually only one stimulus is created (e.g., one simulation, one animation), and these are carefully created to not have internally contradictory information (Ainsworth & Loizou, 2003). Researchers may find very different results from internally consistent sets of multi-modal stimuli that present various aspects of a topic (e.g., the human circulatory system) than from contradictory sets of multi-modal stimuli that present differing interpretations of or hypotheses about a controversial topic.

4. Unique challenges and affordances of multi-source, multi-modal comprehension

A theory of multi-source, multi-modal comprehension needs to identify the unique challenges and affordances that characterize this problem space. Based on major theories from each area (e.g., MD-TRACE, Rouet & Britt, 2011; the DeFT model, Ainsworth, 2006), we would expect the following:

1) A very heavy load on working memory from the different information presented.

Considering even a straightforward learning task such as comparing and contrasting two illustrated texts, we should expect multi-text multi-modal processing to put a very heavy load on working memory. The idea of not overloading WM is very prominent in Cognitive Load Theory (multi-modal learning), which advocates for providing scaffolding to reduce the load on WM (e.g., via segmented worked examples, providing hints, or signaling such as arrows; Paas & Sweller, 2014). Work in multiple text comprehension also acknowledges the heavy load on WM, but usually without suggesting specific solutions (Braasch et al., 2018). Both lines of research typically do not measure WM capacity, it is unclear what type(s) of WM—simple span, complex span, verbal, visuospatial—should be measured, and there is great controversy over how to measure load on WM while learning (de Jong, 2010). There has been some progress in measuring WM load using concurrent task paradigms (e.g., asking the learner to make a mouse click when an irrelevant sound is heard during a narrated video; Park & Brünken, 2015). A theory of multi-source, multi-modal comprehension that was built on measures of WM capacity before and WM load during learning would make important contributions to both the multi-source and multi-modal literature.

2) Difficulty attending to the right part of each stimulus at the right time (e.g., in fast-moving animation, complex diagrams).

With a single multi-modal presentation, it is a challenge for learners to attend to the most task-relevant information within each display (Hegarty, Canham, & Fabrikant, 2010). When information from displays needs to be combined, this may become even more demanding. Imagine a task of learning from two narrated videos about the human circulatory system in sequence. Within each video the learner must focus attention on the portion of the system that the narration is describing. Use of fading, color keys, arrows and other cues only sometimes helps learners to focus on the area(s) being described. Having watched the first video, the learner now has to remember and link information to the second video, while simultaneously attending to the relevant parts of the second video.

Different learning tasks (e.g., watch these videos to be able to write a summary afterwards vs. to give an explanation to a peer afterwards vs. to answer test questions afterwards) make different areas of the stimulus relevant to a different extent (Schnotz & Bannert, 2003). Assigning the task of answering test questions may prompt more memorization of individual parts, whereas assigning the task of giving an explanation may prompt more focus on causal language such as "forceful contraction of the left ventricle is able to send blood the long distances in needs to travel to the rest of the body." Thus, the challenge of attending to the right area(s) of the display is dependent on the specific learning task assigned, as relevance is a task-dependent factor (Hegarty, 2014). A theory of multi-source, multi-modal comprehension needs to account for attending to task-relevant portions of information from these texts.

3) Challenges of remembering the text or diagram source of a piece of information, possibly a memorability advantage for diagram sources.

Multi-text research usually involves presenting passages written from different scientific or historical perspectives, passages which differ in their description and interpretation of a phenomenon (Braasch et al., 2018). By contrast, multi-modal research usually presents combinations of consistent information presented in different modalities, with more or less information overlap between modalities (e.g., diagram-specific, text-specific). It is likely that it is harder to recall which source provided information in a particular proposition when all information comes in text, because visual information can be more memorable. On the other hand, information in animations or videos is quite fleeting, and this may make it harder to recall the source of such information. While sourcing strategies have been very prominent in multi-text research and not in multi-modal research, one could imagine multi-text multi-modal studies that inquire into sourcing from, for example, multiple news video clips showing different interpretations of the same news story. A theory of multi-source, multi-modal comprehension needs to grapple with the role of sourcing when learning from these kinds of displays.

4) Struggle of linking information across "texts" and modalities (within and between texts).

Both multi-text and multi-modal research find a prominent role for within-stimulus inferences (usually called bridging inferences) and inferences that incorporate prior knowledge with to-be-learned information (usually called elaborative inferences; Schmalhofer, McDaniel, & Keefe, 2002). Making more of these inferences characterizes better learners, and can be fostered with various types of strategy instruction and scaffolding (e.g., McNamara, 2004). Nonetheless, these inferences are challenging for learners and in the multi-text literature are found much more often within each text than between texts. These during-learning encoding processes create stronger memory traces when the time comes to retrieve the learned information; that is, better learning. A theory of multi-source, multi-modal comprehension should be able to predict when the inferential load is manageable, for learners at different levels of prior knowledge.

5) Retrieval advantages from encoding information through multiple modalities.

A robust body of evidence from cognitive psychology suggests that encoding the same information in multiple modalities provides an advantage at retrieval, because retrieval cues in each modality or both modalities can result in successful retrieval (Shams & Seitz, 2008). This provides an advantage for learning multi-modal information in multiple texts over print-only information in multiple texts. A theory of multi-source, multi-modal comprehension should incorporate the known retrieval advantages from multi-modal encoding.

6) Challenges of monitoring consistency across different texts and inconsistencies between texts.

Both multi-text and multi-modal research find a prominent role for comparing-and-contrasting across information sources, but multi-text studies focus on monitoring for consistencies and inconsistencies across texts from different sources (perspectives), whereas multi-modal research focuses on comparing-and-contrasting within diagrams (e.g., animal vs. plant cells; Alfieri, Nokes-Malach, & Schunn, 2013). These metacognitive and inferential challenges are likely to be multiplied when learners have to compare-and-contrast within each multimodal text and between multimodal texts. A theory of multi-source, multi-modal comprehension needs to account for the role of monitoring in learning from these displays.

7) Possible biases toward credibility of visual representations.

Despite much research on how manipulative visual displays can be, even 21st century learners may find visuals more credible than print text (Powell, Boomgaarden, De Swert, & de Vreese, 2015). That is, learners may be biased toward the credibility of visuals over text. A theory of multi-source, multi-modal comprehension needs to take into account this possible bias, which may also differ by learner education (e.g., exposure to "media literacy" in classes or courses).

Taking these challenges and affordances into account, scholars developing theories of multi-source, multi-modal comprehension, therefore, have a difficult venture ahead of them, but one that is informed by the rich, thoughtful work of their predecessors in multi-source research and in multi-modal research. I predict that readers with backgrounds in

either multi-source or multi-modal comprehension will find much to learn from all of these papers, and from each of these bodies of literature. I hope that this set of manuscripts lays the foundation for a rich and robust body of work in multi-source, multi-modal comprehension in the years to come, and inspires researchers to expand our knowledge base in this fertile area.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.learninstruc.2018.02.004.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198.

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*(4), 669–681. http://dx.doi.org/10.1016/S0364-0213(03)00033-8.

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113.

Atit, K., Gagnier, K., & Shipley, T. F. (2015). Student gestures aid penetrative thinking. *Journal of Geoscience Education, 63*(1), 66–72. http://dx.doi.org/10.5408/14-008.1.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523–535. http://dx.doi.org/10.1037/0022-0663.96.3.523.

Braasch, J. L. G., Bråten, I., & McCrudden, M. T. (2018). *Handbook of multiple source use.* NY, NY: Routledge.

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

Cook, M., Carter, G., & Wiebe, E. N. (2008). The interpretation of cellular transport graphics by students with low and high prior knowledge. *International Journal of Science Education, 30*(2), 239–261.

Cromley, J. G., Bergey, B. W., Fitzhugh, S. L., Newcombe, N., Wills, T. W., Shipley, T. F., et al. (2013). Effects of three diagram instruction methods on transfer of diagram comprehension skills: The critical role of inference while learning. *Learning and Instruction, 26*, 45–58. http://dx.doi.org/10.1016/j.learninstruc.2013.01.003.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology, 35*, 59–74. http://dx.doi.org/10.1016/j.cedpsych.2009.10.002.

Gegenfurtner, A., Lehtinen, E., & Saljo, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*, 523–552. http://dx.doi.org/10.1007/s10648-011-9174-7.

Hagen, Å. M., Braasch, J. L., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*(S1).

Hegarty, M. (2014). Multimedia learning and the development of mental models. In R. E. Mayer (Ed.). *Cambridge handbook of multimedia learning* (pp. 673–702). (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.033.

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 37–53. http://dx.doi.org/10.1037/a0017683.

Johnson, A. M., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*(2), 97–115.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134.

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.). *Cambridge handbook of multimedia learning.* NY, NY: Cambridge University Press.

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes, 53*(5-6), 430–456.

List, A. Strategies for comprehending and integrating texts and videos. Learning and Instruction (in this issue).

Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly,* 407–432.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.). *Cambridge handbook of multimedia learning.* NY, NY: Cambridge University Press.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011). The effect of diagrams on online reading processes and memory. *Discourse Processes, 48*(2), 69–92.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30.

Paas, F., & Sweller, J. (2014). Cognitive load theory. In R. E. Mayer (Ed.). *Cambridge handbook of multimedia learning.* NY, NY: Cambridge University Press.

Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load—an experimental dual-task study. *Applied Cognitive Psychology, 29*(2), 232–243.

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication, 65*(6), 997–1017.

Rouet, J., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.). *Text relevance and learning from text* (pp. 19–52). Charlotte, NC: Information Age.

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.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*(2), 344–355.

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19*(3), 285–307.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*(2), 105–132.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.). *Cambridge handbook of multimedia learning* (pp. 72–102). (Second ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.006.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13*(2), 141–156.

Schüler, A. The integration of information in a digital, multi-modal learning environment. Learning and Instruction (in this issue).

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417.

Singer Trakhman, L., Alexander, P.A., & Silverman, A. Profiling reading in print and digital mediums. Learning and Instruction (this issue).

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.

van Amselvoord, M. & Schilperoord, J. How number and size of text boxes in argument diagrams affect opinions. Learning and Instruction (this issue).

Van Meter, P. & Cameron, C. The effects of presentation format on multiple document note taking. Learning and Instruction (this issue).

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.