CHAPTER 7

A TECHNOLOGY FOR ASSESSING MULTIPLE SOURCE COMPREHENSION

An Essential Skill of the 21st Century

Susan R. Goldman, Kimberly Lawless, James Pellegrino, and Flori Manning
University of Illinois at Chicago

Jason Braasch
University of Poitiers, France

Kimberley Gomez
University of Pittsburgh

ABSTRACT

Success in today's knowledge society requires the use of multiple sources of information to accomplish personal and professional goals. However, little is known about how young adolescents select, analyze, and synthesize multiple sources to address inquiry tasks. Using evidence-centered design we

Technology-Based Assessments for 21st Century Skills, pages 171–207

> have constructed and tested web-based assessment activities for the selection of useful sources and for the analysis and synthesis of information across text sets. The activities are contextualized in science or history inquiry topics. The assessment activities are designed to provide formative information about student performance that can be used to plan instruction. The design process and the analyses of student performance on these activities are highlighting the unique challenges and opportunities of multiple source comprehension situations.

Two contexts serve to frame the contents of this chapter and its combined focus on technology, literacy, and assessment. The first context is very broad and considers the ways in which educators and researchers should be thinking about technology and its intersection with the processes of teaching, learning, assessment in K–16+ education. To frame these issues we briefly discuss aspects of the vision put forth earlier this year in the National Educational Technology Plan (NETP) (U.S. Department of Education, 2010). The second context is more specific and focuses on a particular set of 21st century competencies that fall under the general heading of digital literacy. The specific focus is on understanding and assessing the knowledge and skills associated with the process of multiple source comprehension. We consider each of these framing contexts in turn as a way to situate the significance of the R & D work described in the remainder of this chapter.

THE NETP VISION FOR TECHNOLOGY AND EDUCATION

The U.S. Congress has mandated the creation of a NETP on roughly a five-year cycle. Prior iterations of the plan have been influential in policymaking at federal, state, and local levels, providing advocacy and guidance for decisions about how funds should be expended and how technology can be employed to best support the nation's education goals. For good reason, earlier plans focused on issues of gaining broad and equitable access to technology. Such an approach made sense as they were developed during an era that saw the rise of the Internet as a public utility, the rapid spread of powerful multimedia computers. The 2010 NETP, however, takes a different stance than the three prior plans because it arrives at a time when access to the Internet is nearly ubiquitous, growing numbers of people carry powerful connected computing devices in their pockets, and our knowledge of how people learn, from both psychological and neuroscience perspectives, is rapidly expanding (e.g., Bransford, Brown, Cocking, Donovan, & Pellegrino, 2000; Sawyer, 2006). Simultaneously, the education system faces unprecedented challenges in terms of student diversity (cognitive, developmental, socioeconomic, linguistic, etc.) and strained budgets.

In this climate, the 2010 NETP differs from its predecessors by focusing on core components of the educational system rather than technology per se.

The majority of the plan addresses current perspectives on *learning*, *assessment*, and *teaching*, as well as needed directions for *future research*. Furthermore, the plan thinks broadly about education, considering learning in both formal and informal contexts, across learners' entire lifetimes. The NETP urges policy makers to focus on grand challenge problems in educational research and development. "Grand challenge problems" are important problems that require bringing together a community of scientists, researchers, practitioners and policy makers to work toward their solution. One of the four grand challenge problems highlighted in the NETP is "*Design and validate an integrated system for designing and implementing valid, reliable, and cost-effective assessments of complex aspects of 21st century expertise and competencies across academic disciplines*" (U.S. Department of Education, 2010, n.p.)

The model of 21st century learning described in the NETP that gives rise to the aforementioned assessment grand challenge requires new and better ways to *measure what matters*; diagnose strengths and weaknesses in the course of learning when there is still time to improve student performance; and involve multiple stakeholders in the process of designing, conducting, and using assessment. In all these activities, technology-based assessments can provide data to drive decisions on the basis of what is best for each and every student and that in aggregate will lead to continuous improvement across our entire education system. Among its recommendations to create more productive assessment systems for education, the NETP advocates for the following:

- Design, develop, and adopt assessments that give students, educators, and other stakeholders timely and actionable feedback about student learning to improve achievement and instructional practices.
- Build the capacity of educators and educational institutions to use technology to improve assessment materials and processes for both formative and summative uses.

For more than a decade, assessment has constituted one of the most controversial issues in education with respect to matters of theory, design, implementation, and educational policy (see e.g., Pellegrino, Chudowsky, & Glaser, 2001). Many of the arguments surround what we assess, how we assess, and the ways in which information derived from assessments is used to shape educational practice. As argued in various sources (e.g., Quellmalz & Pellegrino, 2009), new technologies provide opportunities to shift our assessment systems from a primary focus on summative and accountability practices to one focused instead on formative uses in which assessment in-

formation becomes an integral part of the teaching and learning process. But it is not simply a matter of using technology to shift how we assess students, nor the uses to which we put the information. Most importantly, it is a matter of careful considerations of what can and should become the targets for assessment—the types of knowledge and skill that arise in a 21st century digital world and that are essential for academic and personal success. It is the confluence of method, use, and content that offers the greatest chance for a dramatic shift in the productive integration of assessment into the processes of teaching and learning.

With assessments in place that assess the full range of expertise and competencies reflected in standards, student learning data can be collected and used to continually improve learning outcomes and productivity. For example, such data could be used to create a system of interconnected feedback for students, educators, parents, school leaders, and district administrators. For this to work, relevant data must be made available to the right people at the right time and in the right form. Educators and leaders at all levels of our education system also must be provided with support—tools and training—that can help them manage the assessment process, analyze data, and take appropriate action. This leads us then to the next major topic for consideration—those aspects of literacy that are appropriate and useful in a world dominated by digital content.

DIGITAL LITERACY IN THE 21ST CENTURY

Technological developments in just the first decade of the 21st century have made it impossible to ignore the changing face of literacy. Both the omnipresent access to online information resources and the unprecedented rate at which these resources have grown and continue to grow have focused attention on aspects of literacy beyond those associated with reading traditional print sources: search, evaluation, and integration of information across multiple sources. Although one can argue whether these are truly "new literacies" (New London Group, 1996), it is beyond argument that to be literate in the 21st century means being able to operate in electronic environments to address questions and solve problems that arise in academic, personal, interpersonal, and occupational contexts. Doing so means knowing how to locate relevant sources of information; evaluate sources for reliability, relevance, and quality; analyze the content of each source; and synthesize within and across multiple sources (Coiro, 2009; Coiro & Dobler, 2007; Goldman, 2010; Lawless, Goldman, Gomez, Manning, & Braasch, in press). Furthermore, Web 2.0 with its interactive capabilities increases the likelihood of collaboration and teamwork in carrying out these activities. In sum, technology-based digital environments not only make information

available in multiple formats (e.g., audio and video as well as print) but enable answers and solutions to be produced in a range of formats. Recognition of this reality is beginning to appear in documents such as the recent Common Core Standards for English Language Arts (Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA), 2010).

At the same time, extant data suggest that formal educational settings are providing very limited opportunities to acquire the literacy skills needed for success in the digital age (Goldman et al., ~~in press~~). A few research studies have examined upper-level high school students and college freshmen and reveal that these students use rudimentary approaches to locating, evaluating, and integrating information from multiple sources. For example, students often use imprecise key words in initial searches and have difficulty refining these for more targeted searches (Henry, 2006; Recker, Walker & Lawless, 2003; Wallace, McCory, Kupperman & Kracjik, 2000). In addition, it has been demonstrated that learners' strategies for selecting from among a set of sources prioritize content overlap between the topic of the task and the information source, with limited attention devoted to evaluating the credibility or reliability of the information by examining the author, where the information was published, or the type of publication (Braasch et al., 2009; Britt & Anglinksas, 2002; Coiro & Doubler, 2007; Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). However, other research indicates that sensitivity to the reliability of information sources is associated with better memory and learning of the information (Kim & Millis, 2006; Stadtler & Bromme, 2007; Strømsø, Bråten, & Britt, 2010; Wiley et al., 2009). When it comes to analyzing and synthesizing within and across information sources, the research indicates that learners have difficulty differentiating claims from evidence and conclusions (Brem, Russell, & Weems, 2001; Korpan, Bisanz, Bisanz, & Henderson, 1997; Norris & Phillips, 1994; Phillips & Norris, 1999) and tend to approach the different information sources separately and uncritically (Foltz, Britt, & Perfetti, 1996; Greene, 1993, 1994; Rouet, Britt, Mason, & Perfetti, 1996). When information from multiple sources is present in a response, evidence of conceptual integration is difficult to find (Goldman et al., in press; Mateos & Solé, 2009).

It is important to note, however, that a few recently published studies indicate that both sourcing and integration are amenable to opportunities to learn. For example, several studies demonstrate that providing learners with instructions to evaluate sources and the basis for doing so increase sensitivity to differences in reliability (Britt & Anglinksas, 2002; Stadtler & Bromme, 2007; Wiley et al., 2009). Other studies demonstrate that making the connections among information sources more apparent through graphical overviews has a positive effect on cross-source intertextual connections (Salmerón, Baccino, Cañas, Madrid, & Fajardo, 2009; Salmerón,

Gil, Bråten, & Strømsø, 2010). Typically, however, the impact of such supports is found to interact with prior knowledge and epistemological orientation toward the topic as well as the type of task the learner is attempting to do (Bråten & Strømsø, 2006, 2010; Gil, Braten, Vidal-Abarca & Strømsø, 2010; Lawless, Schrader, & Mayall, 2007).

The foregoing research findings stand in marked contrast to youths' involvement with digital media, largely in out-of-school settings. Indeed, youth are often much more actively engaged with digital media in out-of-school contexts than in in-school contexts (Lenhart, Rainie, & Lewis, 2001; Perez, 2009). The ethnographic studies of Digital Youth conducted over a three-year period by Ito and colleagues (Ito et al., 2009) indicate the existence of a robust youth culture built around online games, social networking sites, and mobile texting and communication devices. And often they are multitasking multiple forms of input at the same time (Lenhart, Hitlin, & Madden, 2005). The rapidity with which digital youth are becoming functionally literate in multimodal and multi-source contexts outside of school emphasizes the gap between what happens in formal educational settings and what happens outside of school. And their mastery of these skills belies the notion that multiple source activities are too complex for middle and high school students. As the gap between in-school and out-of-school literacy practices widens, society runs the risk of creating a youth culture that increasingly views school as irrelevant to what matters outside of school and to future career paths.

It appears then that from a number of perspectives it is important to bring a greater awareness of, and emphasis on, providing instructional support for acquiring the literacies demanded by the technology-based digital world of the 21st century. There are several challenges to address if this is to happen. First, educational practitioners need a better sense of the competencies, knowledge, and skills that need to be supported. Second, they need instructional techniques and materials that can be used to support them. Third, they need to know what knowledge and skills learners bring to the instructional situation and what progress they are making as a result of engaging in instruction. We take the position that high quality assessments, especially those intended to be used formatively, can address the first and third challenges and are thus critical to creating opportunities for youth to engage with 21st century literacies. Formative assessments can define the knowledge and skills that are involved and specify the performances that indicate that they are present. Good formative assessments make clear what students should know and be able to do and what we would take as evidence that they know and can do it (Pellegrino, Chudowsky, & Glaser, 2002). They make this clear to teachers, students, parents, and other stakeholders. This brings us to a consideration of the third section of this chapter: Our work

on the assessment of a specific subset of 21st century literacies: multiple source comprehension.

THE ASSESSMENT OF MULTIPLE SOURCE COMPREHENSION

We approached the definition of 21st century skills in literacy from the standpoint of being able to use information from multiple sources of information to achieve a functional goal, such as answering an inquiry question or solving some problem. Too often comprehension is siloed as a goal in and of itself without practical relevance or connection to the world in which we live. Perhaps that is why over half of the adult population does not read well enough to meet their own health needs (reference). Thus in the assessment context we developed, students are using source texts to accomplish a specific purpose, consistent with the increased emphasis on functional reading (OECD 2002, 2004, 2006).

Our approach to assessing multiple source comprehension relied on the formulation of assessment put forth in the National Research Council report *Knowing what students know* (Pellegrino, Chudowsky, & Glaser, 2002). According to that report, assessment is a process of reasoning from evidence—that is, conclusions about what a student knows or has learned are based on evidence of that knowledge or learning that can be externalized through some sort of observable performance. For quality assessments, three components of the reasoning process have to be aligned: (1) a model of how students represent knowledge and the skills they must appropriate to develop competence; (2) the observations that provide evidence in the form of student performance; and 3) the process for interpreting the evidence with respect to the model of competence. The Evidence-Centered Design approach (Mislevy, Steinberg, & Almond, 2003) provides a systematic way to specify the model, the evidence, and the interpretation and is the conceptual design process we used to develop our assessment of multiple source comprehension.

In embarking on the assessment development process, we first constrained the type of multiple source comprehension that we wanted to assess by focusing on situations in which students were using multiple sources to answer an inquiry question in history or science. Historians and scientists routinely use multiple text sources to address questions in their fields to generate new theories, data, explanations, and knowledge claims (Bazerman, 1985; Berkencotter & Huckin, 1995; Goldman, 2004; Perfetti, Britt, & Georgi, 1995; Shanahan & Shanahan, 2008; VanSledright, 2002; Wineburg, 1991, 1994). In doing so they evaluate the credibility of the information source, the reliability of the data, and the strength of the evidence sup-

porting various claims. However, the importance of each of these processes and the criteria used in the evaluation process are not the same across the two disciplines (Lee & Spratley, 2010; Moje & O'Brien, 2001; Shanahan & Shanahan, 2008). From an instructional viewpoint, both domains involve both information look-up (close-ended questions) and more extended inquiry projects (open-ended questions) (Stahl, Hynd, Britton, McNish & Bosquet, 1996; Wade & Moje, 2002; Williams & Gomez, 2002). Thus, developing multiple source comprehension assessment in these two disciplines allows us to specify a general set of knowledge and skills for engaging with multiple sources as well as realize the specification of the observations and the interpretation of the performance in the context of the practices of the different disciplines.

In the section that follows, we discuss the process we undertook in the development of our assessment across the three components of ECD. We purposively have allocated more discussion to the section detailing the development of the student model, because it is this step that is most often neglected in the development of assessments. That is to say, often researchers and evaluators jump prematurely past the specification of *what* must be assessed into *how* something will be assessed. One of the main purposes of the ECD approach is to prevent jumping to the design of assessment tasks that appear to have "face validity" without specifying in some detail the actual construct to be measured and the forms of evidence that specific tasks need to provide to support inferences about student competence. As we illustrate below, this process of specifying a domain model for purposes of designing valid assessment tasks is both challenging and laborious.

Evidence-Centered Design of Multiple Source Comprehension for Inquiry

The ECD process begins with a clear specification of the knowledge and skills that define competence in the domain of interest. This domain model becomes the basis for developing the student model. The student model is derived by creating claims regarding student performance (e.g., *The student can determine what information is relevant to the inquiry problem)* and the evidence in the student performance that is needed to support the claim. The task model defines the characteristics of the activities in which students will engage and thereby generate observations that speak to specific knowledge and skills in the student model. The interpretive model concerns how to appropriately "fit" the student model and the observations together, that is how to use the observations as evidence for particular claims. In other words, the ECD approach provides conceptual guidance to defining what needs to be assessed, how to assess it, and how to make sense of that which is assessed.

The Domain Model

In developing our multiple source comprehension assessment, we based the domain model on extant theories, analyses, and empirical findings in library and information sciences, discourse comprehension, and literacy practices used by scientists and historians when reading in their fields. This led us to an initial student model illustrated in Figure 7.1. The model consists of five components that constitute groups of knowledge and skills that are relevant to successfully using multiple sources to address inquiry questions.

Interpreting the task is the process whereby the learner comes to understand the objectives, limitations, and boundaries of the task or problem and the kind of information that is relevant to addressing it. It may involve students posing their own question or responding to questions posed by others. One issue is the degree to which learners interpret tasks and adopt task goals that reflect the intended task as conceived of by the task creator. Indeed, students, including college undergraduates, often simplify open-ended, inquiry questions by turning them into close-ended questions for which they seek a single answer, often using a single source (Kuhlthau, 1993; Marchionini, 1995; Wallace, Kupperman, & Krajcik, 2000; Wiley et al., 2009). As such, correct uptake of a task is critical to any inquiry activity.

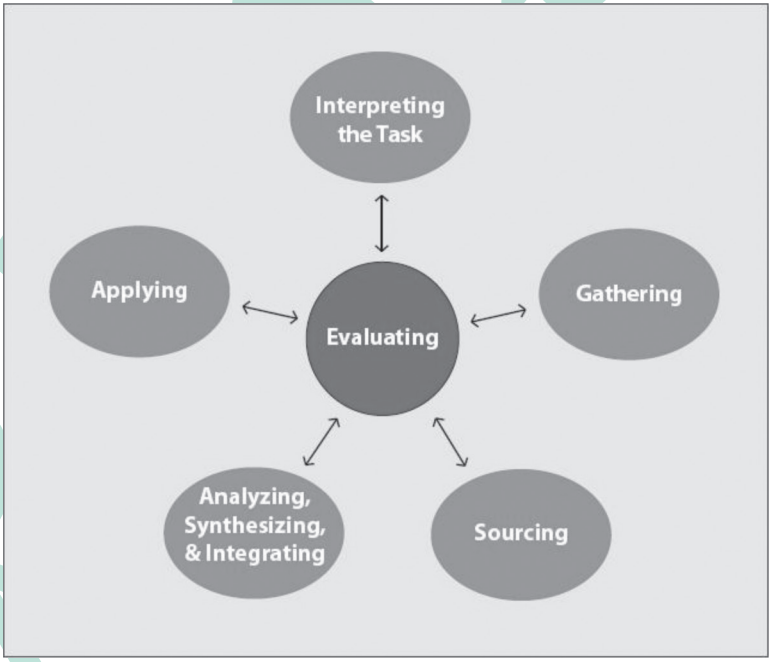


Figure 7.1 A six component multiple source comprehension model.

Gathering refers to processes associated with finding information that can be used to address the task. Learners might brainstorm various possible sources and how to locate them. Efficiency in this process involves a preliminary screening and may contribute to refining, altering, or substituting a new problem or question.

Sourcing and selecting refers to processes associated with determining the usefulness of sources for accomplishing the task based on initial screening of information sources that result from the gathering phase. This phase, ideally, reflects efforts to use information about a source, including its topic (indicated in title or brief summary), who wrote it, and when it was published to determine an initial estimation of its usefulness. It also involves estimating if one has sufficient information to address the problem. Decisions about sourcing and selecting may undergo revision as deeper analysis of the sources occurs.

Analyzing, synthesizing, and integrating information within sources and across multiple sources constitutes the processes of determining what information is in a source and whether it is relevant to the task. Some sources may contain information that is relevant and useful as well as information that is not. The learner has to critically evaluate the content from this perspective. Synthesis refers to determining how information from individual sources relates to the information in other sources; it involves comparison and contrast to determine whether and how information is consistent or conflicting. In the course of analysis, synthesis, and efforts to integrate information to address the task, the learner may determine that more sources are needed or some that looked relevant upon initial screening are not.

Applying information to accomplish the task requires that learners put the information together in a form that meets the constraints of the task. Learners must make decisions about whether the information adequately addresses the problem or question and how to "assemble" the information. Learners' knowledge of the norms and conventions for communicating disciplinary content play an important role in this phase.

Finally, *evaluation* plays a central role throughout multiple source comprehension in that it occurs within each component. In addition, evaluation also serves an executive coordination function by governing movement from one component to another (see Figure 7.1). Within each component, evaluation plays a key role in regulating the processes of that component (e.g., deciding when enough sources have been gathered, or which are relevant). However, the components are also interdependent in that the kinds of sources one gathers might necessitate a re-interpretation of the task, or upon evaluating the relevance of gathered sources, one might determine that more sources need to be gathered. In other words, evaluation occurs within each component and is critical to the coordination among the components. Our use of evaluation is consistent with meta-

cognitive aspects of reading multiple sources (Azevedo & Cromley, 2004; Coiro & Doubler, 2007).

The Student Model

Once the six overarching components in the domain model were identified, we developed the student model by unpacking each of the components in Figure 7.1 through a process of answering the question "What is meant by each component?" in terms of the claims about the knowledge and skills that a "competent" student would possess and therefore be able to demonstrate (Mislevy et al., 2003). This is an iterative process that ends when the claims specify skills and knowledge that are amenable to the development of task models of observable performances that can be used to provide the critical forms of evidence relevant to specific claims. Table 7.1 shows an abridged version of this unpacking for the Sourcing/Selecting

TABLE 7.1 Example of the Unpacking Process for Sourcing/Selection Component with Accompanying Claim and Evidence Statements

Subcomponent	Claim Statement Stem: The student makes use of . . .	Evidence Statement Stem: The work includes information . . .
Relevance	content information during the sourcing process	about the relevance of the content for answering the inquiry question
Reliability	attributes of the source information during the sourcing process	about the importance of source attributes
Source Attributes		
Author	author information in the sourcing process	about the credibility of the author or efforts to determine credibility of the author
Venue	publishing location in the sourcing process	about the credibility of the publication location or efforts to determine where something was published and the credibility of it
Currency/Date	publication relative to the content of the task in the sourcing process	indicating attention to date of publication in relation to task
Type	differences among kinds of resources (e.g., primary vs. secondary; fiction vs. nonfiction; opinion piece/editorial vs. news story) relative to their utility for completing the task	about differences among kinds of resources and their appropriateness for the task

component. The first column lists subcomponents of Sourcing/Selecting, the second illustrates the nature of Claim Statements, and the third illustrates Evidence Statements and their relationship to the claims.

What the subcomponents indicate is that sourcing involves determining the relevance of a source—is the content related to answering the question or solving the problem, as well as the "reliability" of the source. "Reliability" refers to the amount and type of bias or perspective attributable to the information in a source. Various attributes of an information source provide important cues to reliability and are important in guiding how learners interpret and evaluate the information with respect to its usefulness for their task. For example, in history knowing something about the author and the time frame of a document is informative with respect to the perspective that might be reflected in the document. That is, a British officer writing about the Boston Tea Party would be expected to have a different perspective on the event than a citizen living in Boston who was paying the tariff on tea. Journal entries or personal letters describing the event might be expected to reflect these differences in perspective, and a learner might therefore expect differences in the accounts. In science, the date and whether an article is published in a peer-reviewed journal are important to evaluating the quality of the data. Instrumentation available in the 21st century makes the collection of certain kinds of data more precise than would have been possible with the instrumentation 60 years ago. Differences in precision of the data might be very important for explaining discrepancies between data sets. The importance of an attribute in determining usefulness would be expected to fluctuate depending on the discipline and task context. We engaged in a similar process for unpacking all of the components shown in Figure 7.1. The results are depicted in Figure 7.2.

The Task Model

Moving from the claim-evidence student model to actual assessments of components and subcomponents involves developing task models that make it possible to gather the kind of evidence described in the third column of Table 7.1. The task model defines the context in which the "work" is produced as well as other relevant design features of the task situation used to collect the observations. In other words, the task model describes the situations and tasks that could be used to elicit student work that can provide observations that warrant statements about specific skills or knowledge as specified in a claim. A good task model is basically a template that isolates various attributes that can be manipulated to create multiple functionally equivalent tasks. We use the sourcing component as a context for illustrating how a task model is specified.

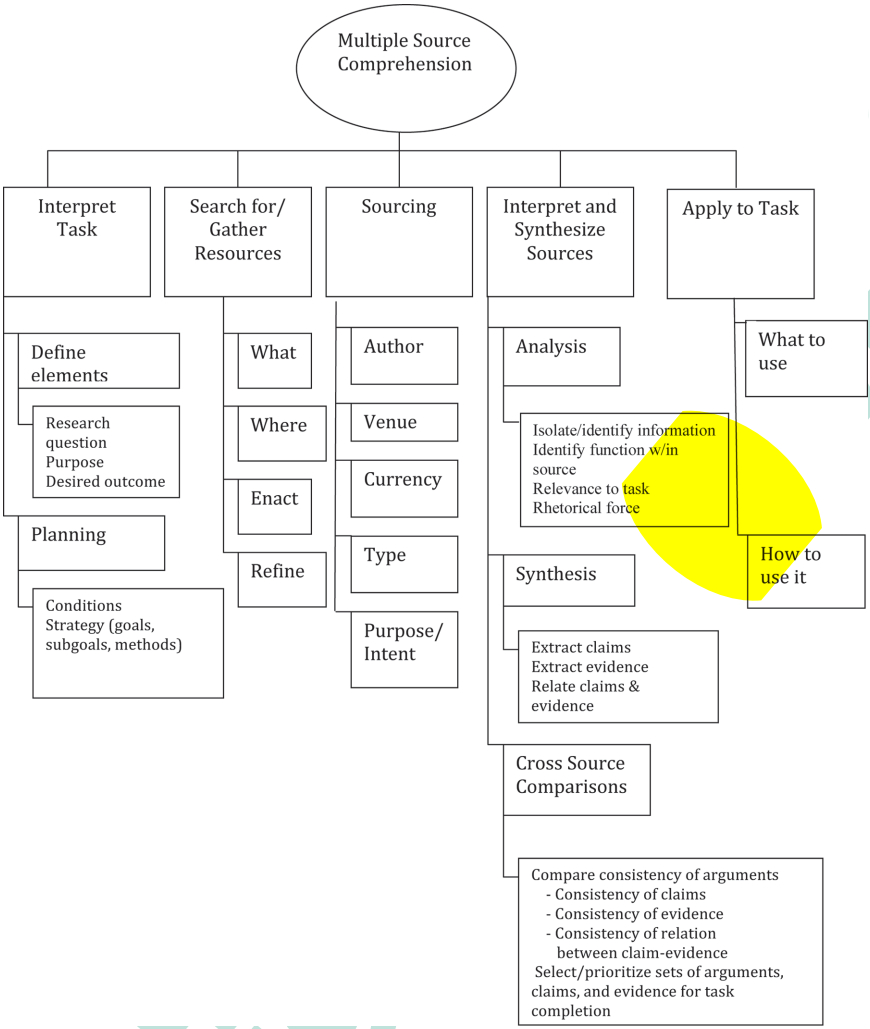


Figure 7.2 The Domain Model components and Student Model subcomponents.

The task model for sourcing represents the structure and features of sourcing task situations in terms of six parameters: 1) source, 2) work product, 3) context, 4) medium, 5) scaffolding, and 6) instructions. The source parameter defines the specific attributes of the texts that need to be specified when designing a sourcing task (e.g., relevance of content, author information, venue of publication, type of publication, etc). The source parameter also delineates some additional considerations that must be taken into account, such as the total number of sources that must be examined.

The work product parameter represents the type of work product students must produce, such as Likert rating, forced-choice yes/no response, essay, and so on. The context parameter indicates the type and amount of contextualization present for a given task and includes attributes such as the presence/absence of cover story in the task. The medium parameter denotes the medium on which a task is performed (e.g., computer or paper and pencil). The scaffolding parameter characterizes the type and degrees of scaffolding provided in the task. Finally, the instruction parameter designates the specific attributes of the instructions provided to students (e.g., where they appear, how much detail is provided and how explicit they are). In this way, the task model makes the process of assessment task design transparent because all the attributes of an assessment task situation are explicitly specified or explicitly marked as unspecified. This process, in turn, constrains the interpretation of the data the assessment task yields.

It was apparent to us that developing task models and actual task situations to assess each of the components shown in Figures 7.1 and 7.2 was more than we were going to be able to do. We decided to specifically focus task model development on just two of the components: Sourcing/Selection and Analysis, Synthesis & Integration. We prioritized these two because they seemed to us to most distinguish single source from multiple source comprehension and to constitute the core knowledge and skills important for multiple source comprehension in an electronic age (Coiro & Doubler, 2007; Goldman et al., 2010).

The Interpretive Model

The interpretive model relates the observations to the student model, indicating what the observations made in an assessment task situation imply for claims about associated knowledge and skills. The interpretive model includes the rules used for scoring or evaluating responses and is shaped by the purpose of the assessment.[1] That is, our interest was in creating a low-stakes assessment situation that would yield information that could help teachers plan multiple source comprehension instruction. As such, the interpretive model was geared toward defining what students *were* doing in the assessment task situation and what that implied about "next steps" in instruction relative to the specific subcomponent tapped by the particular assessment task. We discuss some assessment task situations and how we identified and scored subcomponents in that task in more detail in subsequent sections. Had we been developing a summative assessment for purposes of ranking students, the interpretation of the observed performances would have proceeded differently. The goal of the interpretation model in the context of formative assessment is to provide teachers with information

useful to instructional decision-making. We thus attempted to develop interpretive models that reflected different performance profiles that had implications for instructional focus. For example, as will be elaborated below, if students were using information from only one information source in responding to the inquiry question, that would imply a different instructional strategy than one appropriate for students who were pulling in information from more than one source but doing so in a list format.

In the section of the paper that follows, we illustrate how we created tasks and materials to assess two components of the domain model for multiple source comprehension: Sourcing/Selecting sources and Analysis and Synthesis of Information Sources. We have implemented this process for two history topics (Immigration to Chicago and Civil Rights in the 60s) and two science topics (climate change and threats to the fresh water supply). We use the Immigration to Chicago topic to illustrate implementation.

SOURCING/SELECTING AND ANALYSIS AND SYNTHESIS

The assessment development work in which we have been engaged is designed to provide formative assessment information to teachers of middle school students, beginning with grade five (approximately 11 years old) and extending to grade eight (approximately 14 years). We developed our student model to reflect the competencies of a proficient multiple source comprehender, but our task models defined parameters within which students in this age/grade range might be able to demonstrate mastery. This meant defining the number of sources and difficulty levels of them as well as identifying content that students in this age range would be sufficiently familiar with so that prior knowledge would not severely hamper performance. The topic that is used to illustrate the assessment of sourcing/selection and analysis/synthesis is the movement of people to Chicago between 1830 and 1930, transforming it from a small town of 100 people in 1830 to a metropolis of over 3,000,000 people in 1930. The overall inquiry task for students was to address the question "Why did so many people move to Chicago between 1830 and 1930?"

The Sourcing/Selection Task Situation

In the domain model, the Sourcing component operates on sources that have been returned through an initial searching and gathering process. For example, one might submit a key word or go to a Google search to gather an initial set of sources. The items among the list of sources "returned" vary in their utility for completing the task. Sourcing, as we have defined it in

our domain model, refers to processes of "filtering" the items in the list of sources and selecting a subset for further, more in-depth examination. In our design, the Sourcing/Selection Task situation defines useful sources as those that are relevant to the topic and task and that are trustworthy. The source topic along with the attribute subcomponents shown in Figure 7.2 provide information important to determining relevance and trustworthiness. The task situation for Sourcing consisted of three activities to assess students' skill at each of these: Relevance Judgments, Trustworthiness Judgments, and Usefulness Ranking. Students received a score for each of these activities. The task model specified the basic characteristics of the set of sources: The set of eight needed to reflect a range of relevance and trustworthiness but were modified versions of authentic sources found through Google searches. The information about each source was similar to what might be present in a bibliographic reference and abstract of an article: For each source, there was a title, a 25-word summary describing the content, date and publication venue, author information, and type of publication.

The set of eight sources was rated by an independent set of expert judges (members of the research team not involved in generating the sources but familiar with criteria for reliability in history). According to the expert ratings, two of the eight sources were not relevant, three were highly relevant, and three were somewhat relevant. This basic design of the source set constitutes the task model enabling the replication of the activity for additional inquiry topics.

In the Chicago Immigration task, the students were introduced to the inquiry question "Why did so many people come to Chicago between 1830 and 1930?" and were told they were going to decide on the usefulness of eight sources for answering this question. They were further told that sources were useful if they were relevant to the task and trustworthy. They then completed three activities: Relevance Judgments, Trustworthiness Judgments and Usefulness Rankings.

The first activity—Relevance Judgments—was designed to assess whether or not students could effectively discriminate relevant from irrelevant sources of information based on the content summary and title of each source. In order to help insure that we were assessing students' skills at determining whether a source was relevant and not simply their knowledge of the definition of relevance, the instructions defined relevance in the context of the assessment activity. Figure 7.3 is a screen shot from our web-based selection task in which we define relevance and provide students with two questions to think about in making the judgment. We purposively used the word relevance to make connection with academic language and disciplinary vocabulary. All information was audio recorded so there was voice over for each screen.

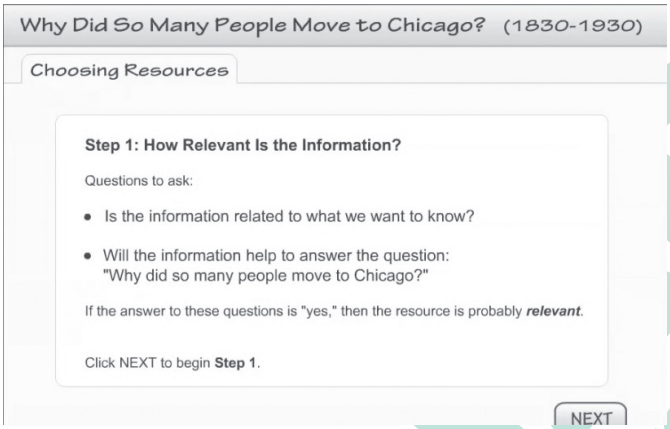


Figure 7.3 Screen shot from assessment application: Criteria for determining Relevance.

Students saw the title of each of the eight sources and clicked on the source to see the summary and make the rating. Using a three-point Likert-type scale (1 = highly relevant, 2 = somewhat relevant, 3 = not relevant) students judged the relevance of the content of eight sources of information in relation to the Chicago inquiry question. A sample screen for this judgment is shown in Figure 7.4.

Once all eight sources had been rated the student was shown the result and had an opportunity to revise any of the relevance judgments before proceeding to the second selection activity, trustworthiness judgment.

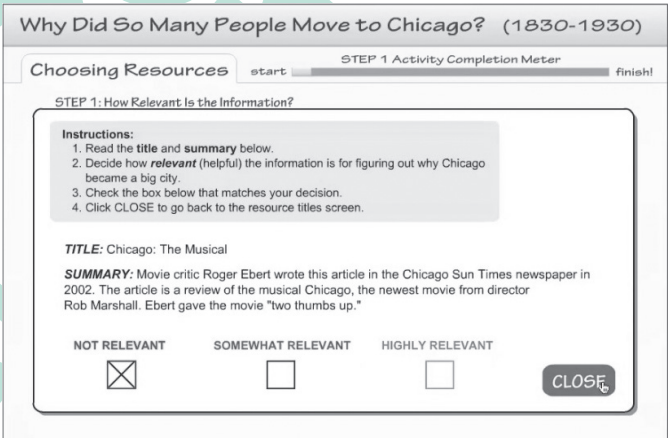


Figure 7.4 Screen shot of Relevance Judgment.

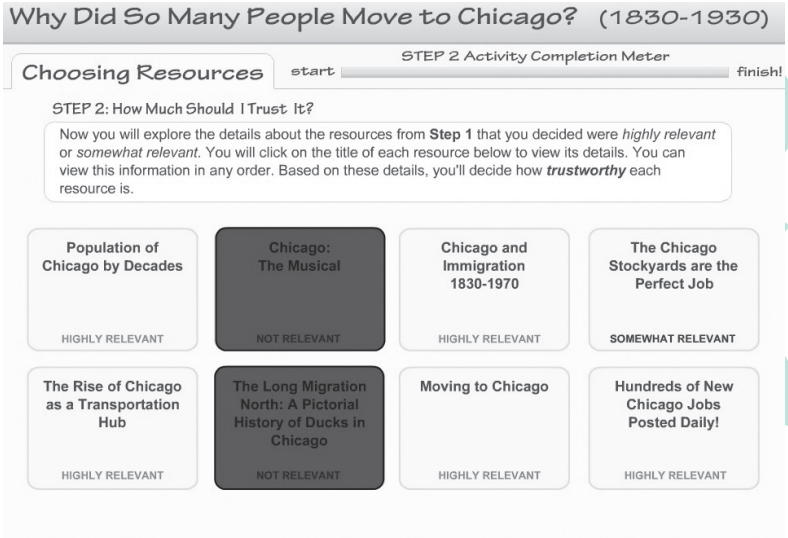
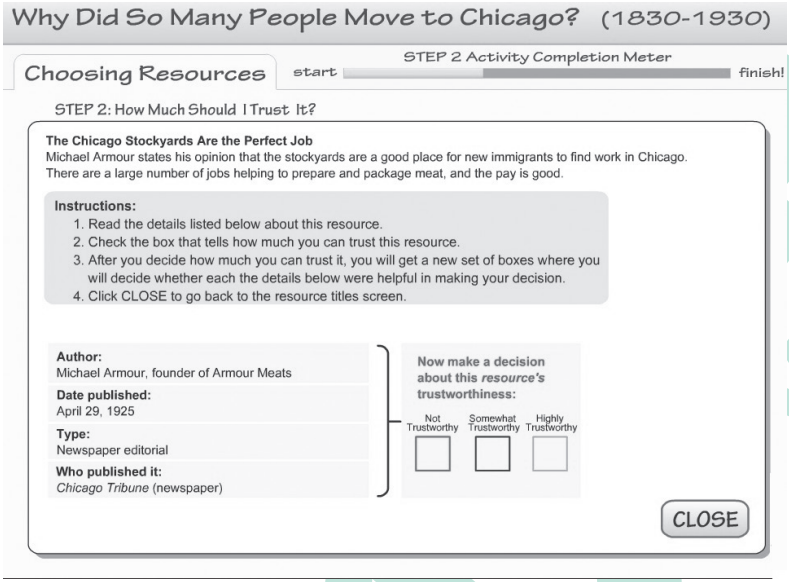


Figure 7.5 Screen shot: Introduction to the Trustworthiness Judgment.

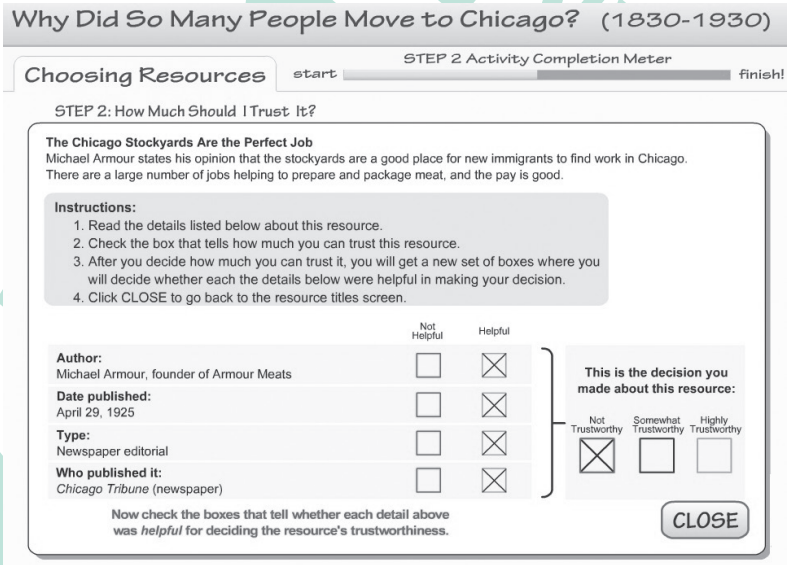
Sources that a student rated as "highly relevant" or "somewhat relevant" were retained for further examination for trustworthiness. Students make no additional judgments about sources that they had deemed "not relevant." Figure 7.5 shows the introduction to the trustworthiness judgment.

In the second activity the students are asked to rate the trustworthiness of any source they judged relevant by considering and rating four source attributes (author, type, publication date, and publication source). Hence this task requires students to think about things such as the credibility of the author, the reliability of the information and its currency. Students need to carefully attend to the source attributes in the context of the content and the inquiry question. For example, a geese migration expert may be highly credible in his/her field, but does not have any apparent expertise in historical trends in human population growth. Students make a holistic rating of trustworthiness and then provide an indication of whether each source attribute was helpful or not in deciding trustworthiness. Again, the holistic rating is three-point, Likert-type scale indicating highly trustworthy, somewhat trustworthy, and not trustworthy. The screens for these decisions are shown in Figure 7.6.

Only sources rated as "highly trustworthy" or "somewhat trustworthy" move forward to the Usefulness ranking activity. For this activity, students are asked to rank order the sources they rated as relevant and trustworthy



(a)



(b)

Figure 7.6 Screen Shots of (a) overall Trustworthiness judgment, and (b) judgment about importance of each attribute to making the judgment.

Figure 7.7 Screen Shot of the Usefulness Ranking activity.

in terms of overall usefulness in answering the inquiry question, "Why did Chicago became a big city?" This task is accomplished by having the student assign a first place "award ribbon" to the source they thought was most useful. The students continued to award 2nd, 3rd, . . . *n*th place ribbons to sources until they had assigned a ribbon to each of their relevant and trustworthy sources (see Figure 7.7).

Elsewhere we have discussed the details of the statistical and analytic techniques that we used to convert the judgment and rating data into scores that allowed us to characterize performance on the usefulness task and how the relevance and trustworthiness judgments related to the overall usefulness decision (Lawless et al., in press). For present purposes, we simply summarize the main findings that have now been replicated over several samples of students across grades five through eight.

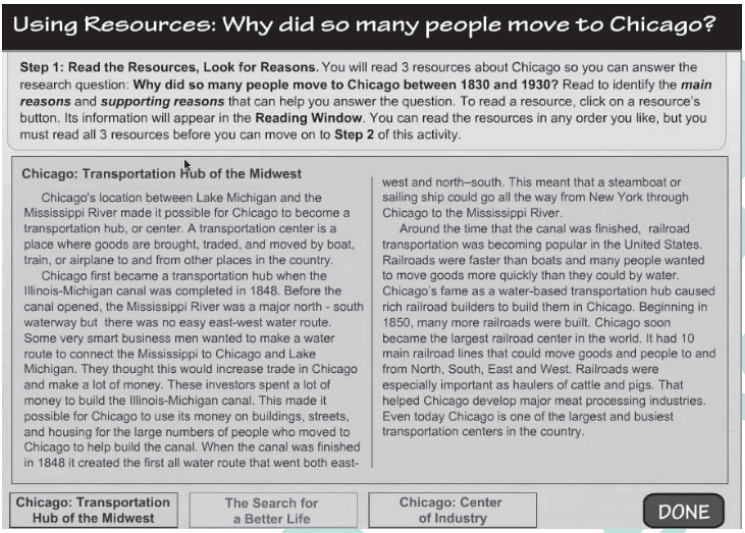
Some students were very good at distinguishing useful from not useful sources, while others were not. For example, in a sample of 64 fifth grade students, 26 students met the criteria for good discrimination of useful from not useful sources, while 23 met the criteria for poor discrimination of sources (Braasch, et al., 2009; Lawless et al., in press). Students who performed at higher levels as compared to those performing at lower levels on the usefulness ranking task also performed at higher levels on the relevance judgment task; however, performance on trustworthiness judgments did not differ significantly. Pilot trials on the trustworthiness task using an open-ended response format replicated the attribute selection data found in the forced-choice format shown in Figure 7.6. The similar pattern across

the two response formats suggests that the results are not an artifact of the response format.

One finding emerging across several samples of fifth to eighth grade students who have participated in the selection task is that those students who perform at higher levels on the usefulness judgments are also more efficient in terms of making use of attribute information. That is, they pay more attention to attribute information for relevant sources than for irrelevant. In contrast, those who perform at lower levels spend time on the attributes for all sources. Thus most students seem to know that source features such as author, date, and so on are important to consider, but students less proficient at this task are more unsure of when or why they are important to consider. Interestingly, when contrasted, the group of students who performed better on this task does not significantly differ on their performance on the district administered, standardized reading achievement test from students who performed poorly on sourcing. Thus, sourcing/selection—in particular the relevance judgment subcomponent of this task—appears to be tapping skills and knowledge not tapped by traditional assessments of reading comprehension.

The Analysis and Synthesis Task Situation

The task model for analysis and synthesis specifies the use of information from three source texts for a complete answer. The content of the texts was based on a consideration of the factors that historians typically discuss to account for the movement of people from place to place. These are *push* factors that encourage people to leave their current location, *pull* factors that draw people to a new location, and *enabling* factors that make it possible for people to get from their current location to new locations. Because we were interested in synthesis of information across texts, the task model specified that each text deal with one of the factors, but include some overlap in vocabulary that would assist learners in seeing the connections across the push, pull, and enabling factors. Thus, one text dealt with *push* factors (Titled: *The Search for a Better Life* and was about the inability of farming to support families), a second with *pull* factors (Titled: *Chicago: Center of Industry* was about industrialization in Chicago and availability of jobs for unskilled labor), and the third with *enabling* factors (Titled: *Chicago: Transportation Hub of the Midwest* was about rail and water transportation enabling the movement of people and goods into and out of Chicago). The text set was constrained to have minimal but some surface level connections in the form of overlapping words. The information and ideas in the three texts complemented rather than contradicted each other.


Using Resources: Why did so many people move to Chicago?

Step 1: Read the Resources, Look for Reasons. You will read 3 resources about Chicago so you can answer the research question: **Why did so many people move to Chicago between 1830 and 1930?** Read to identify the ***main reasons*** and ***supporting reasons*** that can help you answer the question. To read a resource, click on a resource's button. Its information will appear in the **Reading Window**. You can read the resources in any order you like, but you must read all 3 resources before you can move on to **Step 2** of this activity.

Chicago: Transportation Hub of the Midwest

Chicago's location between Lake Michigan and the Mississippi River made it possible for Chicago to become a transportation hub, or center. A transportation center is a place where goods are brought, traded, and moved by boat, train, or airplane to and from other places in the country.

Chicago first became a transportation hub when the Illinois-Michigan canal was completed in 1848. Before the canal opened, the Mississippi River was a major north - south waterway but there was no easy east-west water route. Some very smart business men wanted to make a water route to connect the Mississippi to Chicago and Lake Michigan. They thought this would increase trade in Chicago and make a lot of money. These investors spent a lot of money to build the Illinois-Michigan canal. This made it possible for Chicago to use its money on buildings, streets, and housing for the large numbers of people who moved to Chicago to help build the canal. When the canal was finished in 1848 it created the first all water route that went both east-west and north–south. This meant that a steamboat or sailing ship could go all the way from New York through Chicago to the Mississippi River.

Around the time that the canal was finished, railroad transportation was becoming popular in the United States. Railroads were faster than boats and many people wanted to move goods more quickly than they could by water. Chicago's fame as a water-based transportation hub caused rich railroad builders to build them in Chicago. Beginning in 1850, many more railroads were built. Chicago soon became the largest railroad center in the world. It had 10 main railroad lines that could move goods and people to and from North, South, East and West. Railroads were especially important as haulers of cattle and pigs. That helped Chicago develop major meat processing industries. Even today Chicago is one of the largest and busiest transportation centers in the country.

Chicago: Transportation Hub of the Midwest | The Search for a Better Life | Chicago: Center of Industry | DONE

Figure 7.8 Screenshot of Synthesis Task Reading Instructions.

This task is designed to assess skills at using information from multiple texts to address the inquiry question, and we wanted to ensure that students began with a good understanding that this is what was intended in the task definition. Thus, the assessment itself provided scaffolding that defined using multiple resources as a jigsaw puzzle: The answer comes from many sources and you have to fit the reasons you find together like pieces in a jigsaw puzzle to answer the question. Students were instructed to read each source so they could find the main and supporting reasons to answer the inquiry question. The screenshot in Figure 7.8 shows the reading instructions for the Chicago task and the text conveying the *enabling* factors.

After reading the texts, students moved on to write their responses to the essay question. The writing screen had two sections: one for writing and one for bringing up the individual texts again. Thus, the texts were available to students during the writing portion of the assessment. Students were free to edit their writing. Log files enabled us to capture timing and sequence information during reading and writing portions of the assessment.

The interpretive model for the analysis and synthesis task specified a means of examining the essays for evidence that students: (1) were selectively including or summarizing information from the individual texts consistent with what was most relevant to the inquiry question (analysis); and (2) were connecting the information across the three texts and the relationships that were being captured in these connections (e.g., push-pull; enable movement, etc.) (synthesis). To conduct these examinations of the data, we created content structures for each text that specified which phrases or

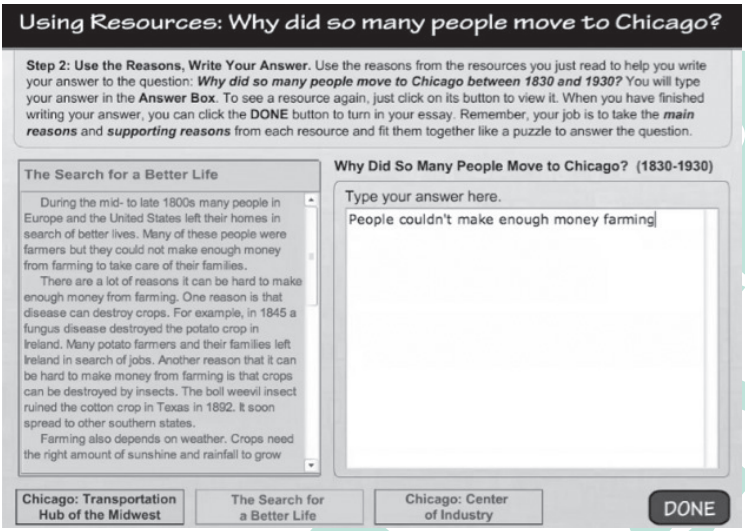


Figure 7.9 Screenshot of the Writing Phase of the Synthesis task.

sentences provided main, supporting, or detail information with respect to the theme/title of the text. A main idea was defined as the main reason(s) that supported the theme of the specific text. In the case of the *pull* factors text, these were statements about the growth of jobs in Chicago due to the emergence of three major industries. Supporting ideas were defined as information that showed that the three industries developed and had needs for people to fill various kinds of jobs. Details were ideas that elaborated the supporting ideas in a variety of ways (e.g., providing specific examples of a more general supporting idea). These content structures are similar to what Kintsch refers to as the *textbase* representation (Kintsch, 1994; Goldman, 2004). Box 7.1 presents the content structure for the *pull* factor text used with the Chicago task.

We mapped the student essay content to the content structure template to look at evidence of selective inclusion of information relevant to the inquiry question. In other words, the content structure mapping is a proxy indicator of students' skill at analyzing individual texts, identifying, and including in their responses the relevant, useful information as opposed to "just any" information from the text. In the case of the Chicago text set, evidence of "better" task-relevant selection would be reflected in a preference for the main ideas and/or supporting ideas over the details in the *push* and *pull* texts but for increased inclusion of details relative to main and supporting ideas in the enabling text. This was so because inquiry task-relevant information in the enabling text was conveyed in what were considered to be details given the theme of that text. Thus, the task model specified that

BOX 7.1: CONTENT STRUCTURE OF *PULL* FACTOR TEXT: CHICAGO: CENTER OF INDUSTRY

Chicago: Center of Industry

In Chicago between 1830 and 1930, businesses grew a lot so they needed people to fill many different kinds of jobs. The jobs were mainly in three kinds of businesses: the railroad industry, meat processing, and retail stores that sold things to people.

The railroad industry provided many kinds of jobs. *Some businesses hired people to build the railroad cars. By the 1850s there were a half dozen companies making railroad cars.* One of the most famous was the Pullman Company, started by George Pullman. He invented the sleeping car. *Another important job was building and repairing the railroad tracks. Other people lifted and moved large crates and machines. All of these jobs needed skills that were not hard to learn* so people from Europe who had recently come to Chicago often filled these jobs. *The engineers who drove the trains had to know a lot more to get hired.*

By 1900, large numbers of people were working in the meat processing industry. *The railroads brought more and more cattle and pigs to Chicago so they could be turned into food and other products. To handle the load, the Union Stockyards opened in 1865. Meat processing factories in large numbers sprung up right next to it. Butchers cut the cattle and pigs into the meat* people bought in grocery stores. *The butchers had special skills. The factories also made other things like soap, glue and shoe polish. Most of these jobs required little skill and were easy to learn. For this reason, many people who came to Chicago without work skills or not speaking English could find work.* But the conditions in the factories were not very pleasant. They smelled bad and were very noisy.

During this same period of time, many retail stores opened and hired salespeople. *For example, Marshall Field opened his department store in 1863 and located it in the middle of downtown. People took jobs selling clothing and household items like sheets and towels. Other stores such as Montgomery Ward and Sears were in Chicago* but they mostly sold things through the mail.

Bold font: Main Ideas; *Italic font: Supporting ideas;* Smaller font size: Details

one text in the set have the property that the thematic content structure needed to be ignored to some degree in a search for information that was most relevant to the inquiry question (see Goldman, 1997 for a discussion of repurposing text information).

Evidence of synthesis required that students make connections across texts using inference processes such as cause-effect, if-then, super set—subset, and related forms of logical reasoning. These connections were not provided in the text set but constitute the construction of a "Situations model" or "Documents model" (Goldman, 2004; Perfetti, Rouet, &

Britt, 1999; Rouet 2006). Students could hypothetically make any number of cross-text, as well as within-text, connections, but only a subset of these might actually be relevant to addressing the inquiry question. To guide our decision making with respect to whether inferences present in the essays should be taken as indicators of cross-text synthesis, we constructed an "expert historian's" "situations model" that expressed the claims and support for them within each text as well as the inferences and connections across texts that expressed logical relations among the push, pull and enabling factors relevant to answering the inquiry question, *Why did so many people come to Chicago between 1830 and 1930?* in the case of the Chicago topic. Thus, the situations model for the Chicago topic shown in Figure 7.10 captures claims and evidence for the push, pull, or enabling factors from each text (the rectangular and circular nodes in Figure 7.10), intra- and intertextual links, and links to the inquiry question. The content of the student essays were mapped against this representation.

We have explored a number of analytic strategies for summarizing the mappings between student essays and the textbase and situations model representations to provide formative information regarding the skills reflected among groups of students. To make claims about analysis, for example, we looked at the percentage of main, supporting, and detail elements

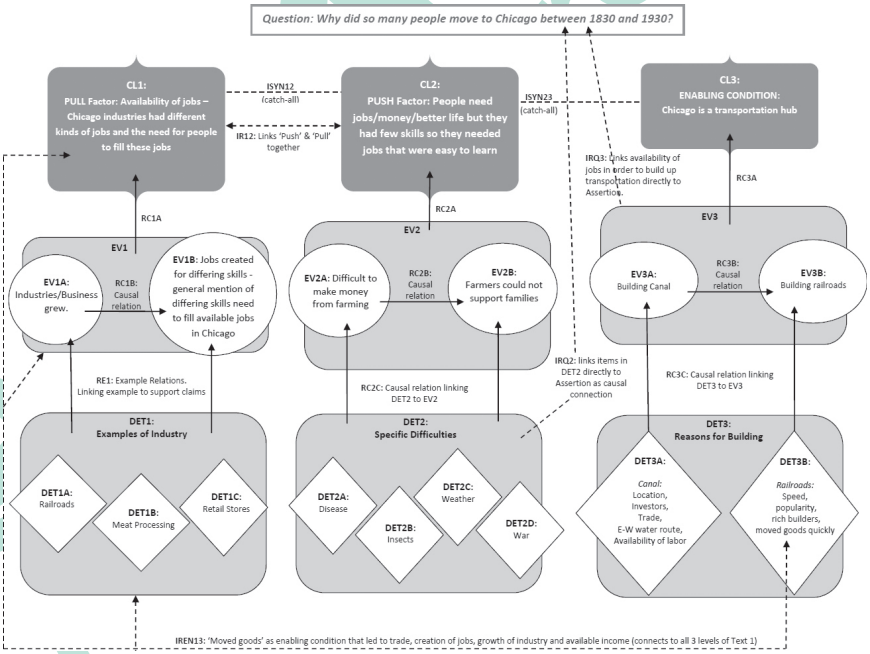


Figure 7.10 Situations Model for the Chicago task.

of each text that were included in the essays of 111 fifth grade students. The data patterns support evidence of analysis: Students were more likely to select main and supporting information from the *pull* or *push* texts than from the *enabling* text but were more likely to select details from the *enabling* text. This pattern is consistent with predictions about relevance of information in the texts based on the content structure and situations model representations. There was also some evidence that students understood that they were to use information from more than a single text to answer the inquiry question: 81% of the students included information from more than one text and 40% included information from all three texts. However, the mappings to the situations model indicated that inferences that reflected true synthesizing links were not that common. The overwhelming tendency was for sequential "blocks" of information from each source to be "listed" in the essay. Box 7.2 provides a contrast between a "blocked" essay and one that reflects greater presence of synthesis. To summarize, the most sophisticated essays reflected selection of the task-relevant information drawn from each of the texts with connecting links that signaled the integration of information across the texts.

BOX 7.2: TWO ESSAYS THAT INCLUDED INFORMATION FROM MULTIPLE TEXT SOURCES BUT REFLECT DIFFERENT DEGREES OF CROSS-TEXT SYNTHESIS.

Example 1: Essay in "blocked" format. There is information from each text in the essay but each is simply enumerated as a reason with no inferential connections across the texts.

A lot of people came to Chicago and now I will write why did so many people come to Chicago.

The first reason I will tell is because of the jobs. Many new busines were opened in Chicago and they needed workers. People were coming from Europe and many other contries to work and earn some money. The second reason is because people were searching an better life. Many of these people were farmers but they could not make enough money from farming to take care of their families. And was a hard time for them. *And third reason is* because of transportation. A transportation center is a place where goods are brought, traded, and moved by boat, train, or airplane to and from other places in the country. Chicago soon became the largest railroad center in the world. It had 10 main railroad lines that could move goods and people to and from north, south, east, and west.

I think those and more other reason's were good to come to Chicago and a lot came here because of these reasons why people came to Chicago. (364)

Example 2: Explicit Connections synthesize push and pull factors

Why did so many people come to Chicago between 1830 and 1930? I believe so many people came to Chicago between 1830–1930 because businesses grew a lot. So because that happened they needed people to fill many different kinds of jobs. All of these jobs needed skills that were not heard to learn. This give people from Europe to come to Chicago to fill these jobs. The(y) was searching for a better life. They just knew how to farm. They did (not) have many other skills. Chicago job had little skills so they went to Chicago. For this little reason, many people who came to Chicago without work skills or not speaking English could find work. `Some of the people came to help build the canal.` **And said the want to stay in Chicago. Before people came to stay in Chicago it was businesses. So the people would be poor and homeless because they would have a job. So it wasn't a lot of people there because it was small businesses. Small businesses could have a lot of people work there.** *So that is why so many people move to Chicago. (54r)*

Example 3: Explicit Connections synthesize push, pull and enabling factors

Why did so many people come to Chicago is because farmers kept losing their crops. Reason one is that disease can destroy crops. For example a fungus disease destroyed the potato crop in Ireland. Another reason is insects can also destroy crops. And because of this farmers a losing a lot of money so they are moving Chicago. *In Chicago between* 1830 and 1930, businesses grew a lot so they needed people to fill many different jobs. Since that happened I think that's why so many people came to Chicago. By 1900, large numbers of people were working in the meat processing industry. `Chicago's location between Lake Michigan and the Mississippi River made it possible for Chicago to become a transportation hub, or center. Chicago first became a transportation hub when the Illinois-Michigan canal was completed in 1848.`

So many people also came to Chicago because there are so many jobs especially the farmers came because their crops were being destroyed so they moved here for better jobs. *That's why people came to Chicago. (309- Explicit Integration)*

Sources and Types of Information in the Essays are indicated by the following different fonts:

Industry text

Better Life text

`Transportation text`

Prior Knowledge

Inferences (any font) are underlined

Inquiry question or rhetorical metastatements

With a larger data set (247 essays across grades five through eight) we were able to consider multiple indicators of synthesis to identify groups of students who performed similarly with respect to their essays. Using a cluster analytic approach, nine variables were submitted as the dependent vector of scores for each student. All of these variables were determined through experimenter-coding of the essays. As this stage, the data were required to be hand coded as we were working on refining our coding rubrics. As we move forward the intent it is automate this coding to be conducted by a computer. Five of the coded variables were elements of the situations model that represented the presence of different kinds of cross-textual connections in students' essays. The other four variables represented: the number of times students linked text content back to the inquiry question, the number of elaboration (non-synthesis) inferences they made, essay word count, and the amount of time spent on the writing portion of the task. Because this selection of input variables contained variables that were both continuous and categorical, a two-step cluster algorithm was performed on the data (Norusis, 2010). To determine which number of possible clusters is most appropriate based on the data, each possible solution is compared using the Schwarz Bayesian Criterion (BIC) as the clustering criterion. The results of the analysis yielded three distinct clusters of students. Based on their profiles across variables, we dubbed the three groups *Satisficers* (50%), *Selectors* (36%), *and Synthesizers* (13%).

- *Satisficers* produced the shortest essays, spent the least time writing, included the least amount of presented information, and did not relate content across texts. Approximately one-third of the essays included information from only one text, another third from two, and the remainder from three.
- *Selectors* produced the longest essays, spent the most time writing, included the most information from the texts, but did not relate content across texts, even though 77% of the essays included information from three texts. None of the two or three text essays included synthesizing inferences. Basically, students copied information from the texts with little to no transformation. Changes reflected selective deletion of information.
- *Synthesizers* connected information across texts, usually relating the push and pull factors. Whether the essays included information from two texts (22%) or from three (77%), all of the essays included at least two synthesizing inferences. Essay length, time spent writing, and amount of information from the presented texts were all lower than those of the Selectors but higher than those of the Satisficers.

As with the selection task, there were no clear relationships between either the text based codes or the cluster membership and performance on standardized reading achievement scores. This finding suggests that skill at answering multiple-choice questions on short passages is not capturing the same knowledge and skills needed to read to answer an inquiry question involving multiple sources of information.

We have used the same task models to develop formative assessments of Sourcing and Analysis/Synthesis for three other topics. We are replicating the basic results that we obtained with the Chicago topic. However, some interesting disciplinary differences are emerging that we think are indicative of the nature of inquiry questions in history and at least certain areas of science. In history, explanations are sought of events that occurred; what is in dispute (typically) is the causes that explain why the event occurred. In contrast, there are many health and environmental sciences issues where the issue is not why a particular event occurred but whether it will occur. This difference has potential implications for the task model and the topic areas in both history and science where parallel models for the two content areas make sense and where they need to be different because the fundamental nature of the inquiry is different.

To summarize our illustration of the application of the ECD process, we illustrated the use of the ECD approach for two particular task situations and demonstrated that they produced a range of performance within and across grades five through eight and shed some light on multiple source comprehension. It is noteworthy that performance on neither the sourcing nor the synthesizing task were significantly related to performance on district-administered reading achievement tests. Thus, the work to date has been very encouraging with respect to elucidating and developing assessments of knowledge and skills that are specific to *multiple* source comprehension.

SUMMARY AND CONCLUSIONS

Our goals for this chapter were threefold: (1) describe elements of a model for multiple source comprehension, (2) illustrate how an evidence-centered design approach could be used productively to develop assessments of student competence, and (3) situate the work in the context of a 21st century web-enabled digital world where technology is a central component of the educative process. We will briefly consider each of these three topic areas in summing up and discussing implications and future directions for the work.

A major premise guiding the lines of inquiry described above is that multiple source comprehension is a significant and challenging aspect of

contemporary literacy in formal educational settings as well as outside of school. In a world where it is commonplace to hear someone say they will "Google it" there is a paucity of information about the efficacy of that process, especially judging whether the results are being used productively to address a significant personal, intellectual or educational question. The evidence that exists on how individuals search for information, how they evaluate what they have found, and how they make sense of and use that information to solve a problem or answer a question suggests that individual skill in engaging in the elements of this complex comprehension process is highly variable. In addition, educators have a ways to go in determining how well students can perform such tasks and in how to assist them. We have begun a line of work that is designed to provide the sorts of formative assessments that can influence instructional decision making and increase the likelihood that instruction will be differentiated to meet students where they are and enhance outcomes through a more sensitive processes of monitoring strengths and weaknesses over time.

We provided two examples of the application of this ECD approach to the production of formative assessments of two important multiple source comprehension components: Sourcing and Analysis/Synthesis. In describing these as well as the other components of the larger multiple source comprehension process, we have attempted to show how the use of an evidence-centered design approach to assessment development is extremely productive since it forces a detailed articulation of a cognitive model of the target domain. Not only does ECD lead to the generation of valid and potentially useful assessments, it is also very useful for theory development. By systematically unpacking the components of the model to identify the claims one wants to make about student competence and the forms of evidence that would provide warrants for those claims, the assumptions about cognition are articulated in a form such that the "constructs" are operationally defined. This can support the design of tasks and a measurement approach where the inferences about student competence are transparent and readily defended. Such a connection among cognition-observation-interpretation is at the heart of the reasoning from an evidence model of assessment described in *Knowing What Students Know* (Pellegrino, Chudowsky & Glaser, 2001).

Nevertheless, there is much more work to be done in that we have taken only the initial steps toward the development of a comprehensive formative assessment of multiple source comprehension. Although we explicated the many types of knowledge and skills that constitute a full model of multiple source comprehension, we have only scratched the surface in investigating but two of the components. The task situations we have used reflect but one task model of each component. Thus we have tapped a fairly limited range of the situations and work products that students might be called on to pro-

duce in multiple source comprehension situations. For example, the text sets were of one particular genre and contained no conflicting information. Other task models for analysis and synthesis might include multiple genres, different points of view on the part of the authors, and contradictions in accounts or explanations. Inferences about students' knowledge and skills relevant to analysis and synthesis under these different task models would be expected to differ from those drawn within the constraints of the task model that we used in this initial work.

It should be clear from the description of the tasks developed and implemented to date for the two different content areas of history and science that there is much work still left to be done in developing assessments. The same can be said of elaboration of the theory of multiple source comprehension. Such work includes extending the development of task models to include other types of texts as well as conducting measurement studies that apply statistical machinery such as generalizability theory (e.g., Shavelson & Webb, 1991) to estimate component sources of variance. Ultimately, the goal is to use the tasks productively in a measurement approach that includes obtaining performance measures that can be used formatively for instructional guidance. Future developments include refinement of the evidence/measurement model as well as automated coding of constructed responses.

We envision carrying out such work under Project READI as part of the Institute of Education Sciences network on *Reading for Understanding* (Goldman, et al., 2009). Thus, it is useful to situate this work in the context of arguments about the nature of reading comprehension and its assessment. For quite some time there has been discontent with the assessment of reading comprehension beyond the most basic levels. This has been especially true in the context of large-scale standardized tests used for summative assessment and accountability purposes. Multiple critiques exist and include: (a) a failure to take into account the structure of the texts to be understood, (b) issues of content knowledge, disciplinary genres, and disciplinary differences in forms of reasoning and argumentation, as well as (c) the limited forms of text processing and comprehension required to answer typical standardized test questions. Perhaps it is not surprising that the recently issued Common Core Standards for English Language Arts take several of these factors into account in describing the nature of the competence expected of students across the three to twelve grade span. There is even allusion to the types of multiple source comprehension skills described in this chapter. Indicative of the disconnect between past work in defining and assessing reading comprehension and the present research is our evidence of the lack of a relationship between scores on standardized reading tests and performance on the components of multiple source comprehension. Such data certainly suggest that the latter are distinct from single source reading as currently measured.

Finally, it is worth situating the research and development effort in the broader NETP context with which we opened this chapter. Two things can be highlighted relative to the goals and objectives spelled out in NETP for assessment and its integration with teaching and learning: (1) *measuring what matters* and (2) developing technology-enabled assessments that can provide instructionally useful information for teachers and students. Our claim is that what we are assessing is one aspect of the competencies that matter in a digital world dominated by the web and ubiquitous access to multiple sources of information known to vary in quality, utility, and truth value. Understanding what components of literacy are required in such a world of information, how to assess them, and how to use that information to assist students in becoming effective comprehenders for educational and everyday uses seems to us an important and worthwhile endeavor in keeping with the goals of NETP. Similarly, we are trying to harness the very same technologies that give rise to the comprehension process to effectively, and seamlessly assess what students are doing in navigating the digital resource world. The goal is to provide that information to their teachers and others as part of a continuous improvement process. In addition to these goals for the assessments we are developing, we see them as a tool for introducing teachers to the kinds of knowledge, skills, and competencies that are required for literacy in the technological age. We have designed them to be relatively seamless with the kinds of tasks teachers might use for instructional purposes. Indeed, the task instructions for the assessments provide the initial steps of potential instructional units designed to develop sourcing or analysis and synthesis of texts and across multiple texts. As such, they might well be embedded within ongoing instructional activities. Thus, while we have a ways to go in attaining the goals described in NETP, hopefully we are at least on the right path.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Yasuhiro Ozuru, Kimberly Richards, Rachel Doherty, Rebecca Penzik, and Kristen Rutkowski in the conduct and analyses of assessment data reported herein. The assessment project described in this article is funded, in part, by the Institute for Education Sciences, U.S. Department of Education (Grant R305G050091). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring organization.

NOTE

1. Our use of the term Interpretive Model most closely matches the evaluation component of the Evidence Model described by Mislevy & Haertel (2006) and is focused on the evidence rules for determining the salient features of student work to be derived from the tasks and that form the basis for claims about student competence. For our present purposes we did not attempt to develop a formal measurement model that is considered by Mislevy & Haertel (2006) as the second major component of the Evidence Model.

REFERENCES

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology, 96*, 523–535.

Baildon, M., & Damico, J. (2009). How do we know?: Students examine issues of credibility with a complicated multimodal Web-based text. *Curriculum Inquiry, 39*(2), 265–285.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication, 2*(1), 3–23.

Braasch, J. L. G., Lawless, K. A., Goldman, S. R., Manning, F., Gomez, K.W., & MacLeod, S. (2009). Evaluating search results: An empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research, 41*(1), 63–82.

Bransford, J. D., Brown, A. L., Cocking, R. R., Donovan, M. S., & Pellegrino, J. W. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press.

Bråten, I., Britt, M. A., Strømsø, H., & Rouet, J. F. (in press). The role of epistemic beliefs in the comprehension of multiple expository texts: Towards an integrated model. *Educational Psychologist.*

Bråten, I., & Strømsø, H. (2006). Effects of personal epistemology on the understanding of multiple texts. *Reading Psychology, 27*, 457–484.

Bråten, I., & Strømsø, H. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes: A Multidisciplinary Journal, 47*(1), 1–31.

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

Britt, M. A., & Aglinskas, C. (2002). Improving student's ability to use source information. *Cognition & Instruction, 20*(40), 485–522.

Coiro, J. (2009). Rethinking Online Reading Assessment. *Educational Leadership, 66*(6), 59–63.

Coiro, J. & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly, 42*(2), 214–257.

Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA). (2010). *The Common Core State Standards for English language*

arts & literacy in history/social studies, science, and technical subjects. Retrieved July, 16, 2010 from http://www.corestandards.org/the-standards/english-language-arts-standards

Cromley, J., & Azevedo, R. (2009). Locating information within extended hypermedia. *Educational Technology Research & Development, 57*(3), 287–313.

Bazerman, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power.* Hillsdale, N. J.: Lawrence Erlbaum Associates.

Foltz, P. W., Britt, M. A., & Perfetti, C. A. (1996). Reasoning from multiple texts: An automatic analysis of readers' situation models. In G. W. Cottrell (Ed.), *Proceedings of the 18th annual Cognitive Science Society* (pp. 110–115). Mahwah, NJ: Erlbaum.

Gil, L., Braten, I., Vidal-Abarca, E., & Stromso, H. I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*(1), 30–68.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris & D. M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 313–347). Greenwich, CT: Information Age Publishing.

Goldman, S. R., et al. (2009). *Reading for understanding across grades 6 through 12: Evidence-based argumentation for disciplinary learning.* Funded July, 2010 by the Institute of Education Sciences, U.S. Department of Education, Grant # R305F100007.

Goldman, S. R., with Lawless, K. A., Gomez, K. W., Braasch, J. L. G., MacLeod, S., & Manning, F. (2010). Literacy in the digital world: Comprehending and learning from multiple sources. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading researchers to life* (pp. 257–284). NY: Guilford.

Goldman, S. R., Ozuru, Y., Braasch, J., Manning, F., Lawless, K. Gomez, K., & Slanovits, M. ~~(in press).~~ Literacies for learning: A multiple source comprehension illustration. ~~To appear~~ in N. L. Stein ~~(Ed.)~~, *Developmental science goes to school: Implications for policy and practice.* New York: Routledge.

Greene, S. (1993). The role of task in the development of academic thinking through reading and writing in a college history course. *Research in the Teaching of English, 27,* 46–75.

Greene, S. (1994). Students as authors in the study of history. In G. Leinhardt, I. Beck, & K. Stainton (Eds.), *Teaching and learning history* (pp. 133–168). Hillsdale, NJ: Lawrence Erlbaum.

Gulli, A., & Signorini, A. (2005, May). *The indexable web is more than 11.5 billion pages.* Paper presented at the 14th International World Wide Web Conference, Chiba, Japan.

Henry, L. A. (2006). SEARCHing for an answer: The critical role of new literacies while reading on the Internet. *The Reading Teacher, 59,* 614–627.

Kim, H. J. J., & Millis, K. (2006). The influence of sourcing and relatedness on event integration. *Discourse Processes, 41*(1), *51–65.*

Korpan, C. A., Bisanz, G. L., Bisanz, J., & Henderson, J. (1997). Assessing scientific literacy: Evaluation of scientific news briefs. *Science Education 81,* 515–532.

~~Kress, G. (2003). *Literacy in the new media age.* London: Routledge.~~

Lawless, K. A., Goldman, S., R., Gomez, K., Manning, F., & Braasch, J. (in press). Assessing multiple source comprehension through Evidence Centered Design. In J. P. Sabatini & E. R. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences.* Lanham, MD: Rowman & Littlefield Publishing.

~~Lawless, K. A., & Schrader, P. G. (2008). Where do we go now? Understanding research on navigation in complex digital worlds. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 267–296) Mahwah, NJ: Lawrence Erlbaum Associates.~~

Lawless, K. A., Schrader, P. G., & Mayall, H. J. (2007). Acquisition of information online: Knowledge, navigation and learning outcomes. *Journal of Literacy Research, 39*(3), 289–306.

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy.* New York, NY: Carnegie Corporation of New York.

Lenhart, A., Hitlin, P. & Madden, M. (2005). Teens and technology: Youth are leading the transition to a fully wired and mobile nation. *Pew Internet & American Life Project.* Retrieved July 10, 2010 from http://www.pewinternet.org/Reports/2005/Teens-and-Technology.aspx

Lenhart, A., Rainie, L., & Lewis, O. (2001). Teenage life online: The rise of the Instant-Message generation and the Internet's impact on friendships and family relationships. *Pew Internet & American Life Project.* Retrieved July 10, 2010 from http://www.pewinternet.org/Reports/2001/Teenage-Life-Online.aspx

Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education—EJPE (Instituto Superior De Psicologia Aplicada), 24*(4), 435–451.

Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. Educational Measurement: Issues and Practice, 25, 6–20.

Mislevy, R. J., Steinberg, L., & Almond, R (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective, 1*, 3–67.

Moje, E. B., & O'Brien, D. G. (Eds.). (2001). *Constructions of literacy: Studies of teaching and learning in and out of secondary classrooms.* Mahwah, NJ: Erlbaum.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66,* 60–92.

Norris, S. P., & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching, 31*, 947–967.

Organization for Economic Co-Operation and Development (OECD). (2002). *Reading for change: Performance and engagement across countries.* Paris: OECD Publications.

Organization for Economic Co-Operation and Development (OECD). (2004). *Learning for tomorrow's world: First results from PISA 2003.* Paris: Author.

Organization for Economic Co-Operation and Development (OECD). (2006). Assessing scientific, reading and mathematical literacy: A framework for PISA 2006. Paris: Author.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perez, S. (2009). Who's online and what are they doing? *Pew Internet and American Life Project.* Retrieved July, 10, 2010 from http://pewinternet.org/Media-Mentions/2009/Whos-Online-and-What-Are-They-Doing-There.aspx

Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-based learning and reasoning: Studies in history.* Hillsdale, NJ: Erlbaum.

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

Phillips, L. M., & Norris, S. P. (1999). Interpreting popular reports of science: what happens when the reader's world meets the world on paper? *International Journal of Science Education, 21,* 317–327.

Quellmalz, E., & Pellegrino, J. W. (2009). Technology and testing. *Science, 323,* 75–79.

Recker, M., Walker, A., & Lawless, K. (2003). What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instructional Science, 31*(4–5), 299–316.

Rouet, J. F. (2006). *The skills of document use: From text comprehension to web-based learning.* Mahwah, NJ: Erlbaum.

Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88,* 478–493.

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

Salmerón, L., Baccino, T., Cañas, J. J., Madrid R. I., & Fajardo, I. (2009). Do graphical overviews facilitate or hinder comprehension in hypertext?, *Computers & Education 53*, 1308–1319.

Salmeron, L., Gil, L., Bråten, I., & Strømsø, H. (2010). Comprehension effects of signaling relationships between documents in search engines. *Computers in Human Behavior, 26*(3), 419–426.

Sanchez, E., & Garcia, J. R. (2009). The relation of knowledge of textual integration devices to expository text comprehension under different assessment conditions. *Reading and Writing: An Interdisciplinary Journal, 22*(9), 1081–1108.

Sawyer, R. K. (Editor). (2006). *Cambridge handbook of the learning sciences.* New York: Cambridge University Press.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage Publications.

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

Stahl, S. A., Hynd, C. R., Britton, B., McNish, M., & Bosquet, D. (1996). What happens when student read multiple source documents in History? *Reading Research Quarterly, 31,* 430–456.

Strømsø, H. I., Bråten, I., & Britt, M. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*(3), 192–204.

U.S. Department of Education. (2010). *Transforming American education: Learning powered by technology.* Washington, DC: U.S. Department of Education. Available online at http://www.ed.gov/technology/netp-2010

VanSledright, B. A. (2002). Confronting history's interpretive paradox while teaching fifth graders to investigate the past. *American Educational Research Journal, 39*, 1089–1115.

Wade, S. E., & Moje, E. B. (2000). The role of text in classroom learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 609–627). Mahwah, NJ: Erlbaum.

Wallace, R. M., Kupperman, J., & Krajcik, J. (2000). Science on the Web: Students online in a sixth-grade classroom. *Journal of the Learning Sciences, 9,* 75–105.

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.

Williams, K. & Gomez, L. (2002). Presumptive literacies in technology-integrated science curriculum. In G. Stahl (Ed.), *Computer support for collaborative learning: foundations for a CSCL community* (pp. 599–600*).* Mahwah, NJ: Lawrence Erlbaum.

Wineburg, S. S. (1994). The cognitive representation of historical texts. In G. Leinhardt, I. L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 85–135). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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*, 73–87.

~~Zhang, S. & Duke, N. K. (2008). Strategies for Internet reading with different reading purposes: A descriptive study of twelve good internet readers. *Journal of Literacy Research, 40*(1), 128–162~~.

Author Queries:

1 – On p. 3 you cite Wallace, McCory, Kupperman & Kracjik, 2000, but this is not included in your references. Please add this item to your reference list. (Is this supposed to be the same as Wallace, Kupperman, Krajcik, 2000 as listed in your references?)

2 – On p. 4 you cite Ito, et al., 2009, but this is not included in your references. Please add this item to your reference list.

3 – On pp. 4 and 5 you cite Pellegrino, Chudowsky, & Glaser, 2002, but this is not included in your references. Please add this item to your reference list. (Should this be Pellegrino, Chudowsky, & Glaser, 2001?)

4 – On p. 5 you cite Wade & Moje, 2002, but this is not included in your references. Please add this item to your reference list. (Should this be Wade & Moje, 2000?)

5 – On p. 6 you cite Kuhlthau, 1993, but this is not included in your references. Please add this item to your reference list.

6 – On p. 6 you cite Marchionini, 1995, but this is not included in your references. Please add this item to your reference list.
7 – On p. 13 you cite Kintsch, 1994, but this is not included in your references. Please add this item to your reference list.
8 – On p. 13 you cite Goldman, 1997, but this is not included in your references. Please add this item to your reference list.
9 – On p. 14 you cite Norris, 2010, but this is not included in your references. Please add this item to your reference list.
10 – You include Baildon, & Damico, 2009 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
11 – You include Bråten, Britt, Strømsø, & Rouet (in press) in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
12 – You include Cromley, & Azevedo, 2009 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
13 – You include Gulli, & Signorini, 2005 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
14 – You include Kress, 2003 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
15 – You include Lawless, & Schrader, 2008 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
16 – You include Sanchez, & Garcia, 2009 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
17 – (See query 4) You include Wade, & Moje, 2000 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list. (Or, if these two items are one and the same, please correct whichever date is incorrect.)
18 – You include Zhang, & Duke, 2008 in your references, but this is not cited in the text. Please add a citation or remove this item from your reference list.
19 – On p. 5 you write "Perhaps that is why over half of the adult population does not read well enough to meet their own health needs (reference)." Was that a note to yourself to add a reference there? Please add a citation and add the reference item to your references.