Literacy demands of 21st-century society require that we move beyond simpler views of reading comprehension in our theories, research, and assessment (Goldman, in press; Goldman et al., 2010, 2012; Lawless, Goldman, Gomez, Manning, & Braasch, in press; Lee & Spratley, 2010; Moje, 2008). For example, electronic technologies make information ubiquitous and increase the likelihood that “the public” of all ages finds itself attempting to bring together multiple sources of information in the course of their personal, academic, and professional lives (Britt, Rouet, & Braasch, Chapter 11, this volume; Coiro, 2009; Coiro & Dobler, 2007; Goldman, 2004; Goldman et al., 2012; Wiley et al., 2009). To “bring together” information from multiple text sources, readers need to go beyond decoding and comprehension processes associated with word—and sentence—level processes, often referred to as simple views of reading (cf. Andrews & Reynolds, Chapter 5, this volume; Landi, Chapter 2, this volume; Oakhill, Cain, McCarthy, & Nightingale, Chapter 7, this volume; Van Dyke & Shankweiler, Chapter 8, this volume). They have to move to interpretive levels and decide what the text means in the context of their purpose or goal. They need to decide what information is relevant; if it fits together coherently; if there are gaps or inconsistencies; and if it contradicts previously acquired knowledge. If such issues arise, readers need to decide what to do about them. The decisions and what to do about them often depend on information about the text, such as who wrote it, when, and for what purpose. (See, for discussion, Bråten, Strømsø, & Britt, 2009; Britt et al., this volume; Britt & Rouet, 2011, 2012; Goldman et al., 2012; Strømsø, Bråten, & Britt, 2010.) These interconnections between reading and reasoning are evident in the following hypothetical example.
A news story hits the networks about an earthquake in China. There has been an intense tremor and major damage to certain areas of the country. Concerned for the welfare of family and friends who live in China, you search the Internet to get additional information, entering *China earthquake*. Your search returns several sources of information: Newspapers releasing flash updates, twitter feeds from individuals in the immediate vicinity, video of the quake captured on cell phones, and links to stories published after prior earthquakes of like magnitude and occurring in similar geographic areas. Across these “texts” a reader must sort out the relevant details. Where was the epicenter? What is the population in that area? What kind of damage can such a quake do? What is the early damage assessment? Are there human injuries? What happens in the aftermath of such a disaster? This information, of course, also needs to be integrated with one’s prior knowledge of major earthquakes, such as those that have occurred in California, Haiti, and Japan. In addition you have to consider the type of source and its author to decide on its reliability and trustworthiness. If the author was someone from the U.S. Geographic Survey, you might be more inclined to accept the claims after a superficial reading of the article than if the author was someone who stood to gain financially from disaster relief efforts. In the latter case, you might actually read the article very closely before deciding on the validity of the information.

The general point is that how people read texts depends on how they reason about the information; and how people reason about the information depends on how people read texts. Reading and reasoning are necessary to arrive at interpretations of the information.

Although going beyond what text says to interpretations of what text means requires entertaining complex conceptualizations of reading comprehension, simpler views have dominated approaches to reading comprehension theories, instruction, and assessment. One might ask why. One response is found in Charles Perfetti’s 2009 keynote address introducing the U.S. Department of Education Institute of Education Sciences’ Reading for Understanding Initiative. During this talk, Perfetti referred to the “conceptual muddle” that results when reading comprehension takes on other aspects of cognition:

*We can expect the comprehension of written language to approximate the comprehension of spoken language. When that happens, then reading comprehension has developed, for practical purposes, to its limiting or asymptotic level. … All other limitations are imposed by linguistic abilities, relevant knowledge, and general intelligence. If we make things more complex than this, we push onto the concept of reading comprehension all these*
other important aspects of cognition, with the muddle that results from conceptual conflation.

(Perfetti, Landi, & Oakhill, 2005; emphasis added)

The challenge then, in moving beyond simple views is maintaining clarity regarding the construct. In this chapter, we illustrate our efforts to maintain conceptual clarity for the construct multiple source comprehension by using the Evidence Centered Design approach to assessment development (ECD; Mislevy, Steinberg, & Almond, 2003).

ECD is particularly appealing as an approach to assessment of complex comprehension as reflected in multiple source comprehension because it begins with the definition of the construct in terms of knowledge, skills, and competencies. Assessment tasks are developed based on explicit statements, called claims, about the knowledge and skills that constitute the construct and the work that would provide evidence of students’ competence with respect to the claims. As such, the ECD process makes clear what students should know and be able to do and what would be taken as evidence that they know and can do it (Pellegrino, Chudowsky, & Glaser, 2001). ECD can be contrasted with a commonly used technique in cognitive science, rational cognitive task analysis (Clark & Estes, 1996). Task analysis begins with the task and asks what knowledge and skills a student would need to successfully complete the task. However, the relation of tasks to constructs is not a central concern.

We provide an overview of ECD and then instantiate it in the context of our efforts to develop assessments of multiple source comprehension in early adolescent students, roughly 10-14 years old. We designed the assessments to serve a formative function by providing teachers with information about their students’ knowledge and skills for engaging in deep understanding of multiple information sources. Engaging in the ECD process took us from a very broad definition of multiple source comprehension to specific operational definitions of two subcomponents of the broader construct: (1) selection of sources useful to answering an inquiry question; and (2) information integration from a set of constrained sources to answer an inquiry question.

Evidence Centered Design: A Brief Overview

The ECD process begins with specifying the knowledge and skills that define competence in the domain, much like models of expert performance in other areas of cognitive sciences. The domain model is the basis for developing the student model, i.e., the knowledge and skills that would define student, as opposed to expert, competence. The student model is expressed as a series of claim statements regarding student performance. They begin with the stem The student can . . . <show, differentiate, determine>. Claim statements cannot use verbs like understand, think about, comprehend. For each claim statement, there is at least
one corresponding evidence statement, using the stem *The work provides information that the student <differentiates, shows, identifies>*. The task model defines the characteristics of the activities in which students will engage and thereby generate observations that fulfill the terms of the evidence statements. It includes the specification of the work product or performance that serves as the vehicle for collecting observations. The evidence model sits between the student and task models and describes how the observations in the work product or performance are scored, evaluated, and interpreted with respect to the claim statements in the student model. Using the ECD approach for assessment development is an iterative process wherein the results of initial tests are used to refine and elaborate domain, task, and evidence models.

**Multiple Source Reading Comprehension: Domain and Student Models**

We developed the construct of multiple source reading comprehension for the functional goal of answering inquiry questions using more than a single text-based resource (Goldman et al., 2010; Lawless et al., in press). We postulated a domain model drawing on extant research and theory in library and information sciences, discourse comprehension, and literacy practices within the disciplines (e.g., Goldman, 2004; Lee & Spratley, 2010; Moje & O’Brien, 2001; Shanahan & Shanahan, 2008; Stadtler & Bromme, 2007).

**The Domain Model**

The domain model includes six components: interpreting the task; gathering resources; sourcing and selecting resources; analyzing, synthesizing, and integrating information within and across sources; applying information to accomplish the task; and evaluating processes and products. We arrived at these components and their description by repeatedly asking the ECD question: *What do we mean by multiple source reading comprehension, as informed by extant research and theories?* (Goldman et al., 2011).

*Interpreting the task* reflects the need to understand the objectives and boundaries of the task, including the kind of question or problem that is posed, the various subproblems that might be involved, and generally the kinds of sources that might be helpful to accomplish the task.

*Gathering* resources refers to finding, identifying, and locating information to address the task.

*Sourcing and selecting* involves whittling down gathered resources to find the most useful sources for the task. Task relevance and reliability are two critical aspects of sourcing and selecting. Decisions about relevance depend on topical relationships between the task and the information source (Braasch et al., 2009). Individuals with prior knowledge of the inquiry topic are likely to be in a better
position to make efficient sourcing and selection decisions than individuals with little knowledge of the task domain. Reliability is influenced by who created the piece, why, and when (Braasch et al., 2009; Bråten et al., 2009; Stromso et al., 2010; Wiley et al., 2009; Wineburg, 1991; Zhang & Duke, 2011). How those dimensions are weighted depends on readers' knowledge of the content area and the field more generally. Evaluation of the results of sourcing and selecting might lead to decisions to go back and gather more resources, perhaps with more refined search criteria.

*Analysis, synthesis, and integration* are closer to traditional conceptions of reading comprehension than are many of the other components of multiple source reading comprehension. They involve basic reading processes for figuring out what the text says, including word recognition, lexical access, parsing, and extracting meaning elements. These processes contribute to the formation of a textbase-level representation—essentially the ideas that are directly and explicitly communicated through the words and their organization in the text (Goldman, Varma, & Coté, 1996; Kintsch, 1988, 1998; van den Broek, Risden, & Husebye-Hartman, 1995; van Dijk & Kintsch, 1983). Theories of single text comprehension also posit that readers create a situation model level of the text that reflects the reader's interpretation of the world referred to by the text. The situation model representation synthesizes information within the text and with prior knowledge. Thus, even single text comprehension involves synthesis in terms of making inferences that are activated by the explicit text and that connect parts of the explicit text and/or the explicit text with prior knowledge (Goldman, Graesser, & van den Broek, 1999; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; McNamara & Kintsch, 1996; Wiley et al., 2009; Zwaan & Radvansky, 1998). These inferences can reflect a variety of types of reasoning, including associative, causal, and other logical relations. In the context of an inquiry task, analysis involves determining the relevance of information to the inquiry question. This is because even though a text as a whole may be topically related to the inquiry, not all of the information in the text may be useful for addressing the inquiry task. Readers have to critically evaluate what is and is not relevant to their inquiry task and integrate that information so it addresses the task.

In multiple source comprehension synthesis and integration operate across sources as well as within single sources and involve at least two additional levels of representation. First, the source level represents information about the sources (e.g., author(s), date of publication, publication venue). Second, the integrated model level reflects as complete a model as the comprehender can get of how the information in the multiple sources fits together with respect to the inquiry task. The integrated model or the integrated mental model (Britt & Rouet, 2012; Britt, et al., this volume) reflects bringing together situation model and textbase levels from individual sources and is akin to the documents model as discussed by Perfetti, Britt, Rouet, and their colleagues (Britt, Perfetti, Sandak,
& Rouet, 1999; Goldman, 2004; Perfetti, Britt, & Georgi, 1995; Perfetti, Rouet, & Britt, 1999; Rouet, 2006). Synthesis across multiple texts involves inferential reasoning processes that include comparison and contrast of information to determine whether and how information is related (e.g., complementary, overlapping, or redundant) or unrelated. And if it is related, whether it is consistent or inconsistent and the type of logical relations that apply. Cross-text synthesis is important in history for corroboration and in science for establishing the reliability of results. Inconsistencies can be of a variety of types with different implications for reading behaviors and interpretations. For instance, to return to the earthquake example, one newsfeed might indicate that there were no fatalities but another might report that several people had died. The contradiction between the two articles might lead readers to evaluate the evidence for the contradictory assertions in the two articles and potentially seek additional information to clarify. Integration involves organizing the outcomes of analysis and synthesis processes to form the integrated model.

Synthesis, as well as integration, are likely to be more difficult in multiple source situations than if the same content was included in a single source because the writer of a single source that contained conflicting information would very likely indicate this conflict through the use of rhetorical devices and signals (e.g., On the other hand, An alternative view, In contrast . . .). In multiple text situations, readers must supply these relational indicators. Furthermore, multiple text situations provide more opportunities for connections and readers need to figure out which are germane in the inquiry context and which are not.

Application involves organizing the information that has been synthesized and integrated so that it complies with the task requirements. This phase involves knowledge of the forms of communication that are appropriate for accomplishing different purposes and the conventions that govern how they are organized. As such, application often involves reading and writing skills. However, we include it in our model of multiple source comprehension because the process of applying the information to accomplish the task constraints may well lead to deeper levels of understanding of the information and perhaps reinterpretation of information and sources that readers thought they already understood sufficiently.

Evaluation operates within each component and also coordinates movement from one component to another, depending on the outcome of attempts within each component. Within each component evaluation regulates processing (e.g., deciding whether a particular source has enough relevant information in it to be useful). The components are also interdependent in that readers may have to revisit task interpretation based on the information in the sources that are found. There may be points or issues that readers determine simply cannot be addressed because the relevant data is simply not available. Or the reader may decide to search further for the relevant data even if it is not in the sources thus far considered. Thus, evaluation is at work when readers decide that the sources they
have selected are insufficient to accomplish the task or when they decide that
information from one source matches up with information from another. Evaluation as conceptualized in this domain model is consistent with discussions of
the metacognitive aspects of reading multiple sources (Azevedo & Cromley,

From this domain model we developed student, task, and evidence models
for two components: sourcing and selecting; and analysis, synthesis, and integra-
tion. In the remainder of this chapter we focus on the analysis, synthesis, and
integration component. Information on the sourcing and selecting component
is available in Braasch et al. (2009), Goldman et al. (2012), and Lawless et al. (in
press). We illustrate how we operationally defined the analysis, synthesis, integra-
tion component for multiple source reading comprehension through the
claim–evidence statements in the student model and the assessment situation
specification in the task model. The scoring and analytic strategies used to evalu-
ate student work products reflect our instantiation of the evidence model.

**Student Model for Analysis, Synthesis, and Integration**

The student model comprises claim–evidence statements for the constructs
analysis, synthesis, and integration as we operationalized them for this multiple
source reading comprehension assessment. The claim–evidence statements indi-
cate what students know and can do based on the evidence in work they
produce. Abridged versions of the specification of these three constructs and
claim–evidence statements for them are provided in Table 12.1. As with the
components of the domain model we arrived at these descriptions by repeatedly
answering the question “What is meant by <subcomponent>?” (Mislevy et al.,
2003) for analysis, synthesis, and integration. The answers are described (column
1) and then expressed in claim–evidence statement (columns 2 and 3). The
claim–evidence statements specify the skills involved in multiple source com-
prehension with respect to inferences about what students can do (the claims)
based on specific features of work products (evidence).

Specifically, the analysis subcomponent involves skills for (1) determining
task relevance of information in a text, and (2) identifying the claim(s) and evid-
ence supporting it (them) within a text. “Reading” the claim–evidence state-
ment for the first skill in Table 12.1, we would make the claim that a student
can show what parts of a text are useful for a task (claim statement) if that
student produces a work product that differentiates between useful and less
useful parts of a text. As discussed below, the Evidence Model specifies how this
differentiation translates into observables, given a particular assessment situation.
The descriptors for synthesis shown in Table 12.1 refer to skills for (1) compar-
ing claims across sources and (2) evidence across sources. The descriptors for the
integration subcomponent show that it involves (1) organizing claims and evid-
ence from different sources so that they can be used to address the inquiry task.
<table>
<thead>
<tr>
<th>Subcomponent</th>
<th>Student model</th>
<th>Evidence model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Claim statement</strong></td>
<td>The student can...</td>
<td>Evidence statement</td>
</tr>
</tbody>
</table>

### Analysis
1. Determine relevance of information to task.
2. Identify claim(s) and evidence in each text.

1. show that, how or why, when or where part of text is useful for task.
2. differentiate between claims and their supporting information.

### Synthesis
1. Compare claims across sources for consistency and relevance to inquiry task.
2. Compare evidence from difference sources.

1. determine which claims agree, disagree, or complement one another.
2. determine which evidence is consistent, which inconsistent across sources.

### Integration
1. Combine similar claims; organize complementary claims.
2. Relate evidence to claims, regardless of how they were introduced in the texts.

1. take claims from different sources and combine them into one claim or set of claims.
2. connect evidence to appropriate claims regardless of original text source.

---

**Evidence from essay**

- Selective inclusion of claim and evidence nodes over details. Absence of distortions of text information.

**Dependent measure**

- Proportion of essay reflecting textbase, especially elements central to inquiry question.

- Inclusion of information from multiple texts, especially multiple claim and/or evidence nodes.

- Mean number of texts reflected in essay content.

- Inclusion of inferences that connect claims from different texts; and evidence within and across texts.

- Mean number of inferences reflected in essay content, especially cross-text inferences.
As specified in (2), this may sometimes involve using evidence presented in support of one claim to support a different claim.

**The Task Model**

The claim–evidence statements in and of themselves tell us little about the topics, inquiry questions, and kinds of texts (genres and complexity levels) in which students produce work products. That is the job of the task model. It describes the assessment situation in terms of the conditions for collecting observations so that the work products are capable of providing the evidence specified in the evidence statements in the student model. These “conditions” include the topic, type of response, type and number of texts, task instructions, timing, and so on. The task model we developed reflects our selection of a particular form of analysis, synthesis, and integration that pilot testing indicated would be developmentally appropriate for students in the grade band 5–8, age 10–14.

In developing the task model, we realized there were a number of different types of synthesis situations. Consider just three types:

1. Synthesizing different text genres to answer an inquiry question (e.g., primary versus secondary documents in history; newspaper accounts of research studies versus peer-reviewed journal reports of experiments);
2. Synthesizing/comparing multiple texts that agree with one another versus contradict one another (e.g., two editorials that take similar positions as compared to two that take opposite sides on a topic such as immigration policies of the U.S.);
3. Synthesis of texts that each contribute a piece of the answer, but do not provide a complete solution to the inquiry task on their own (e.g., an economic analysis in one text and a political analysis in another of fluctuations in the cost of gasoline).

To create assessment situations for each of these requires three different task models. We formulated the task model for the third type of analysis/synthesis/integration, in part because it reflected a beginning level of analysis and synthesis: to answer the inquiry question completely, relevant information would have to be selected from several texts (analysis), claims and supporting information compared and coordinated across texts (synthesis), and organized logically to address the inquiry question (integration). This assessment situation would therefore be able to provide evidence for only the first skill shown in Table 12.1 for each of the subcomponents.

The inquiry task was specified as using the information in a set of texts that were provided to write an essay in response to the question *Why were the civil rights events of 1955–1965 more successful than previous civil rights events?* The text
set contained three texts that presented complementary information regarding
the historical event. One described the federal laws (LAW) that were passed
between 1860 and 1955 and reasons they had limited impact on the rights of
African Americans. The second described direct action efforts (DA) that prior to
1955 had consisted of largely local, small-scale events. The third text described
the advent of television (TV) as a common household item that brought the
news of the day into people’s homes in ways that had not been possible previ-
ously. Intentionally, the genre of each of the texts is similar to “textbook” so
that we could look at analysis, synthesis, and integration independent of issues
of sourcing. Each text contained claims about the role of laws, direct action, or
the advent of television along with evidence to support the claims in the form
of generalizations with specific example events. For example, from the DA text:
generalization: Small groups fought for change in their own communities; specific
events: In the 1930s, Blacks boycotted the Woolworth’s drug stores in Chicago. They
refused to buy anything there because Woolworth’s would not hire Blacks. The three
texts are provided in Table 12.2. The three texts as a set met the parameter in
the task model of providing information that was sufficient for students with
little or no prior knowledge to demonstrate analysis, synthesis, and integration
in an essay answering the inquiry question. Additional parameters in the task
model set conditions on length of the texts, readability, and degree and type of
content overlap among the texts in the text set (e.g., surface overlap of some
words, repetition of events) (see Goldman et al., 2012, for details).

The task model also captures the administration conditions of the assessment
situation. In the present context we report results from an administration of the
civil rights assessment situation to 211 students in grades 5 (n = 70), 6 (n = 90),
and 7 (n = 51). The assessment task was developed as an online activity pro-
grammed in Flash, with an archival database collecting information on student
responses and interactions within the activity. An introductory audio and ani-
mation sequence contextualized the inquiry question for the students and
modeled how to answer a “why” question by searching for the main reasons and
supporting reasons across the available texts that addressed the inquiry question.
Students were told that “these reasons all fit together like [pieces of] a puzzle:
when you put the pieces together, you see the whole picture—the answer to
the question!” As a reminder, the written inquiry question was visible across the
top of the computer screen throughout the duration of the entire task. All
instructions were provided in both text and audio formats.

Students were told they would have three texts to use to answer the inquiry
question in a written essay. The activity was divided into two phases: reading
and writing. To increase the probability that students read the three texts in
their entirety before writing their responses, the students were first asked to read
the three passages (Step 1: reading) for the main and supporting reasons and
could not go on to writing until they had met the minimum “screen time” for
each text (2 minutes per text). By clicking on the titles of each of the three texts
TABLE 12.2 Three texts in the text set for the civil rights inquiry question Why was the civil rights events of 1955–1965 more successful than previous civil rights events?

Legal Rights as a Force for Change

Early civil rights battles were fought in the courtroom in order to change federal law. A law is a rule that is created by an authority figure to protect people, their property and their rights. While many of these battles were successful and created new laws that supported the civil rights of black people, they had little impact on how black people were treated. People had to decide to obey the new laws. But many people did not want to obey them. People not wanting to obey new laws is something that is quite common. Cell phone use while driving is a good example. There are now laws against it but people don’t see why they have to obey that law and still talk or text on their phones while driving a car.

This same thing happened with civil rights laws, some of which date back to the civil war. For example the Civil Rights Act of 1875 created a federal law that gave black people the right to use the same public places as white people, including public restrooms, water fountains, and restaurants. But many people in the Southern states like Alabama, Mississippi, and Georgia did not want black people to use the same facilities as white people. Some lawmakers in these states created new laws that pretended to obey the federal law. They were called “separate but equal” laws. Separate drinking fountains and bathrooms were designated for use by blacks, but they were often in poor working condition compared to those designated for whites. Schools that whites attended had better books and buildings than schools that black students had to go to. All people could ride the same city bus but blacks had to sit in the back while whites could sit wherever they wanted. So while the Civil Rights Act of 1875 changed the federal law, it did not really improve the lives of blacks in many U.S. cities and towns.

In the end, it became clear that making and changing federal law alone was not enough to make real improvements in the treatment and lives of black Americans. Something would need to happen. Black people had to demand their rights. One way of doing this is through direct action. Direct action is a strategy for change where groups of individuals do things that call attention to problems that need to be fixed. Direct action was one way black people fought for their civil rights and were able to make a difference in how they were treated.

Early direct action efforts were small, and made local changes. Small groups fought for change in their own communities. In the 1930s, blacks boycotted the Woolworth’s drug stores in Chicago. They refused to use the public toilets because they were forced to use the bathroom in the back. The boycott lasted 4 months and ended when Woolworth’s finally hired 21 black workers. The Brown v. Board of Education decision in 1954 was another example. Blacks refused to use the public schools because they were forced to ride in the back. The boycott lasted just 4 days and ended after the bus company agreed to let blacks share the middle section of the bus with whites, but not the front. These small victories showed how it was possible for people to come together and make a change within their communities. But many, more people in the country still needed help.

Later direct action events were much bigger, highly organized, and made national changes. Led by the civil rights organizations, many of these events won some big civil rights victories. For example, the National Association for the Advancement of Colored People (NAACP), Dr. Martin Luther King Jr., and the Montgomery Improvement Association (MIA) all worked together to plan and run the Montgomery Bus Boycott (1955–1956). Having all of these groups work together was important because it meant that people in different places could communicate with each other and plan much bigger direct action events that lasted longer. The Montgomery Bus Boycott lasted more than a year and more than 20,000 people participated. As a result, laws changed so that everyone could sit wherever they wanted on the buses. The 1963 March on Washington was another large direct action event planned by many of the major civil rights organizations. At this gathering of over 200,000 people, Dr. Martin Luther King Jr. delivered his famous “I Have a Dream” speech. This event was also televised live as it happened. People all over the country watched on television.

Many people, including the President of the U.S., began to push for equal rights for all American citizens because of the March on Washington.

All of these direct action events were successful because of the way people and organizations worked together. The early, smaller efforts made local changes. The later, larger and more organized events brought civil rights issues to the attention of people across the nation and helped obtain equal rights for all U.S. citizens.

Taking Action and Working Together for Change

Throughout history, people have had to stand up for their rights. One way of doing this is through direct action. Direct action is a strategy for change where groups of individuals do things that call attention to problems that need to be fixed. Direct action was one way black people fought for their civil rights and were able to make a difference in how they were treated.

Early direct action efforts were small, and made local changes. Small groups fought for change in their own communities. In the 1930s, blacks boycotted the Woolworth’s drug stores in Chicago. They refused to use the public toilets because they were forced to use the bathroom in the back. The boycott lasted 4 months and ended when Woolworth’s finally hired 21 black workers. The Brown v. Board of Education decision in 1954 was another example. Blacks refused to use the public schools because they were forced to ride in the back. The boycott lasted just 4 days and ended after the bus company agreed to let blacks share the middle section of the bus with whites, but not the front. These small victories showed how it was possible for people to come together and make a change within their communities. But many, more people in the country still needed help.

Later direct action events were much bigger, highly organized, and made national changes. Led by the civil rights organizations, many of these events won some big civil rights victories. For example, the National Association for the Advancement of Colored People (NAACP), Dr. Martin Luther King Jr., and the Montgomery Improvement Association (MIA) all worked together to plan and run the Montgomery Bus Boycott (1955–1956). Having all of these groups work together was important because it meant that people in different places could communicate with each other and plan much bigger direct action events that lasted longer. The Montgomery Bus Boycott lasted more than a year and more than 20,000 people participated. As a result, laws changed so that everyone could sit wherever they wanted on the buses. The 1963 March on Washington was another large direct action event planned by many of the major civil rights organizations. At this gathering of over 200,000 people, Dr. Martin Luther King Jr. delivered his famous “I Have a Dream” speech. This event was also televised live as it happened. People all over the country watched on television.

Many people, including the President of the U.S., began to push for equal rights for all American citizens because of the March on Washington.

All of these direct action events were successful because of the way people and organizations worked together. The early, smaller efforts made local changes. The later, larger and more organized events brought civil rights issues to the attention of people across the nation and helped obtain equal rights for all U.S. citizens.

The Rise of Television

For many of us it is hard to imagine a world without television, but in the early 1950s, only 1 in 10 U.S. households had a television set. By the end of the 1950s, the number of American homes with a television jumped to nearly 90%. Television had an advantage over other forms of mass media like radio. People getting their news from the radio could only listen to the events, but with a television people could see and listen. As a result, many people felt more connected to television events.

Television provided people with a way to experience important news events as they occurred. Even all over the world were broadcast into people’s living rooms. People were introduced to a whole variety of things that, prior to television, they could only imagine. Everyone watching television news or shows shared a common experience. Rich and poor people of all ages experienced televised events at the same time. In 1959, for example, Americans all over the country watched the televised event of the very first time an American was sent into outer space.

Television also helped shape public opinion by giving people the change to see the world through the eyes of others. People began to understand more about political issues because they could actually see, on television, the problems other people had to face. In a sense, people could actually experience those faraway events for themselves. For example, people watching television in the northern United States could see for themselves the unfair living conditions endured by blacks living in the South. They also saw how blacks struggled against these conditions. In the 1950s and 1960s, television viewers saw news broadcasts of angry white mobs trying to keep black students out of white schools, attacks by police and dogs on children and civil rights protesters, and bus boycotts. It was one thing to hear on the radio that protesters were being sprayed by fire hoses but seeing it on television made a more powerful impression. Viewers witnessed with their own eyes that the protesters could not even stand up against the force of the water as it was sprayed on them.

Because these television broadcasts were national, a lot more people became aware of the poor treatment of blacks in the South. This helped change public opinion and resulted in many people deciding to join in the fight for equal rights for all U.S. Citizens. It also demonstrates the important role television has played in the shaping of society as we know it today.
(whose placement order on the screen was randomized), students were able to view and read each text, at any pace and order of texts. Students spent approximately 10–15 minutes reading all three of the texts. Students then proceeded to Step 2: writing. To write their essay, they were told to use the three texts (still available to them) for reference while typing out their responses. Note that although the texts were available, they could not scroll over text and cut and paste it. (See Goldman et al., 2012, for further information and screen shots of the software application.) Students spent approximately 15–30 minutes on the writing portion of the task.

**The Evidence Model**

The evidence model specifies how the observations, in this case the contents of the essays, are to be evaluated with respect to the claim–evidence statements in the student model. Our a priori evidence model is provided in column 4, Table 12.1. We anticipated that evidence of the information relevance skill (analysis) would be reflected in selective inclusion of main reasons and key supporting evidence or details from the LAW and DA text regarding why actions prior to 1955 were less successful than those occurring between 1955 and 1965. We also expected inclusion of information about the role of television in raising awareness of civil rights issues. The key evidence for synthesis was inclusion of information from more than a single text source. For integration it was inferences that connected claims and/or evidence across texts. Column 5 of Table 12.1 provides the dependent measure(s) used as indicators of the various types of evidence described in column 4.

To enable us to systematically map essay contents to an integrated model that could be constructed from the three texts to address the inquiry question, we developed the schematic shown in Figure 12.1. Content from each of the texts is represented as a series of vertically organized content nodes, starting at the top with the overall main reason from each text. These are claim nodes in Figure 12.1 and are labeled CL1, CL2, and CL3 for LAW, DA, and TV texts respectively. Beneath each claim are the evidence nodes (EV#) supporting the claims. At the third level are nodes reflecting details of specific events (DET) that illustrate the more general evidence statements. Connections between nodes that were provided in the texts themselves (labeled R) are shown in Figure 12.1 as solid lines. Inferred, between-text connections (labeled IREN##) are indicated with dotted lines. There were also unanticipated cross-text inferential connections present in students’ essays. These are shown as ISYN in Figure 12.1. We indexed specific sentences from the texts to each node and link. The integrated model served as a template for scoring the information contained in student essays.

According to the integrated model in Figure 12.1, an essay that provided a complete answer to the inquiry question would make three claims: the spread
of television was a key enabling condition in the success of the civil rights movement in the 1955–1965 period (CL3) because (IREN23) it made more people aware of direct action efforts (CL2) and the inadequacy of federal laws to extend civil rights to African Americans (CL1). With respect to the student model, the “best” essay would be able to provide evidence of synthesis and integration in terms of inferences connecting elements across texts, indicated by IREN and ISYN links.

Students’ essays were scored in two passes, the first with respect to the presented texts (texbase coding) and second with respect to the integrated model (Figure 12.1). In the first pass, raters worked from essays that had been parsed into idea units. Each idea unit was “mapped” to the texts and was coded to indicate what text and which sentence it was drawn from, as well as whether it was a direct, verbatim copy of the text information or whether it was a paraphrased version of the text information. This texbase coding allowed us to identify the depth and breadth of content coverage within each text and across the text set included in each student’s essay. As such, we were able to examine if students were extracting the important information from the texts relative to the inquiry question and if they were able to successfully differentiate the relevant from the irrelevant sections of the original texts. The texbase codes thus
served as a proxy measure for the analysis component of the student model. In addition to the textbase codes, essays were coded for the presence of prior knowledge, extratextual inferences, and distortions (misinterpretations) of the text base.

In the second scoring pass, coders examined student essays for the elements (nodes and links) of the integrated model (see Figure 12.1). These data allowed us to understand which components of the text set students were “synthesizing” in order to develop a coherent response to the inquiry question. Each essay was coded by a minimum of two independent coders who obtained an interrater reliability of at least 85%, with differences being resolved in conference.

During the coding process and through initial analyses examining the dependent measures shown in Table 12.1, it became clear that simply looking at each dependent measure independent of what else was in the essay provided an incomplete picture of performance on the subcomponents. We decided to explore cluster analysis as a means of revealing patterns across a set of dependent measures. We derived six measures of what was included in the essays as related to the integrated model (number of essay statements that were direct copies of statements from LAW, DA, or TV; number of essay statements that were paraphrases of statements from LAW, DA, or TV) and four additional indicators of essay content: essay word count, inferences not related to synthesis, instances of prior knowledge use, and the number of distortions of presented content. These variables were entered into a two-step cluster analysis as the dependent vector of scores for each student. A unique feature of the two-step algorithm is the automatic selection of the number of clusters versus a prespecified cluster amount that is entered by the researcher as is called for by other clustering methods (Norusis, 2010).

The analysis yielded three clusters across all students (5th–7th grade). Essays within each resulting cluster were examined with respect to the type and proportion of included content: text-based information, distortions of the text-based information, inferences, and prior knowledge. In addition, the amount of information that was “copy and pasted” versus transformed was considered. Based on the trends in these variables exhibited by each of the different clusters, we labeled them replicators (20.9%, n = 44), transformers (31.3%, n = 66), and minimalists (47.8%, n = 101). We then looked at indicators of analysis, synthesis, and integration for each of these groups.

Replicators wrote essays for which the majority of content (58%) consisted of information that was lifted directly from the texts, untransformed. A smaller percentage of their text-based content (22%) was paraphrased, with distortions of text-based information (14%) rounding out the major sources of content in their written work. These students’ essays contained very few, if any, inferences (2%) or prior knowledge (4%). Replicators spent the most time on the writing task (M = 1687 seconds) and wrote the longest essays (M = 196.5 words).

Transformers’ essays were characterized by a very large proportion (76%) of paraphrased text-based content, in contrast to a relatively small amount of
copied-and-pasted content from the texts (10%). These students’ essays also contained few distortions of text-based information (6%), almost no instances of prior knowledge (1%), and the highest proportion of inferences (7%). It appears that both the replicators and the transformers made the inclusion of text-based information a priority in their work, but the transformers provided a more sophisticated treatment of this content in their written responses. The transformers spent less time writing than the replicators, an average of 1294 seconds, and the average word count was 179.8.

Minimalists, unlike the replicators and transformers, provided work that was characterized by larger proportions of non-text-based content (inferences: 5%; prior knowledge: 11%; distortions of information from the texts: 18%). Only 66% of minimalists’ essays consisted of text-based content, in comparison with essays by the replicators (80%) and transformers (86%). The minimalists spent the least time writing their responses to the inquiry question (M = 927 seconds) and, not surprisingly, produced the shortest essays (M = 97.8 words).

Evidence of synthesis and integration skills was differentially reflected in the three clusters. The transformers demonstrated the strongest evidence of synthesis: 92% of their essays included content from two or three texts, overall M = 2.42 texts included. In contrast, only 40% of the minimalists included content from more than one text (M = 1.43). Of the replicators, 75% included information from more than one text (M = 2.11). Integration, indicated by the presence of inferences that connected content nodes across texts, occurred at very low levels in the minimalists and the replicators (15% and 16% of essays, respectively). On the other hand, 45% of the transformers’ essays reflected integration in the form of cross-text inferences. The majority of these inferences connected the details across the LAW and DA texts or made miscellaneous connections at the general level across the TV and DA texts.

Our analyses of skill at determining relevant information (analysis) are interesting in that the evidence level (EV in Figure 12.1) of the integrated model was skipped over by most of the students and claim statements were included at moderate levels but largely for the DA text. Details from each of the texts were the most frequent nodes included in the essays, with the specific patterns quite similar for replicators and transformers. Minimalists included almost no information from the TV or the LAW texts but showed a pattern of content inclusion similar to the other two clusters for the DA text. These patterns are complex and we continue to analyze them but they clearly do not accord with the a priori expectations of the Evidence model. What they do reflect is a not surprising tendency for students in this age range to focus on concrete, single-instance events as exemplars of general claims.

It is interesting to note that higher quantities of time and/or writing did not necessarily produce higher-quality work products. Although the replicators wrote lengthier essays, their strategy of copying phrases directly from the texts in formulating their responses was not as effective as the transformers’ method of
paraphrasing (and synthesizing) this information in order convey what they thought was important to say. The quick-and-short approach adopted by the minimalists is also telling with respect to the quality of their responses, which reflected less emphasis on the inclusion of text-based information and instead favored the incorporation of proportionately more instances of prior knowledge in the construction of their essays.

Conclusions and Continuing Issues

Synthesizing and integrating information stemming from multiple, separate sources of information are clearly comprehension skills for which systematic instruction is needed. In our sample of over 200 children in grades 5–7, approximately 30 students demonstrated evidence of inferring connections across information presented in at least two different texts. Furthermore, these connections were at the detail level of specific events, with selection biased toward those details with relatively high semantic overlap across texts. The inclusion of claim statements in essays was infrequent and when evidence was used to support claims it tended to be through specific example events rather than generalizations over several examples. In other words, analysis for relevance favored the instantiations of generalized statements rather than the generalized statements, expressed as topic sentences in the texts.

The ECD process played a key role in allowing us to make more transparent precisely what performances constitute complex or deep reading comprehension, leading to limited but clearly defined constructs for analysis, synthesis, and integration. It is in fact rather humbling to realize that the grain size at which we specified these three subcomponents was much more constrained than we had initially anticipated. This grain size resulted from the ECD process of specifying the knowledge and skills along with the kind of evidence that would indicate that such knowledge and skills were present. While measured at a more microlevel than anticipated, the assessment process allowed us to identify distinct profiles and approaches to analysis, synthesis, and integration. The clarity of the construct assessed provides an excellent basis for designing instruction that begins where students are and moves them forward along a developmental trajectory. As such, the assessment development process and the resulting assessment situations and task performance examples can be educative for teachers regarding aspects of multiple source comprehension that might be appropriate foci for different groups of students.

Thus, our initial efforts to assess multiple source comprehension, even with the limited way in which we operationalized analysis, synthesis, and integration, indicate clear starting points for efforts to support students’ developing complex comprehension skills. Indeed, in the context of an ongoing collaborative effort, Project READI, we are developing and researching instructional modules specific to one of three disciplinary areas (literature, history, or science). The
modules are designed to support multiple source comprehension in service of addressing inquiry questions appropriate for that content area. Instructional routines for engaging students in close reading of text material are a key component of these modules. “Close reading” is framed in terms of criteria for determining relevance of information within a source (analysis), connections within and across sources (synthesis), and practices of argumentation (integration) appropriate to each of the disciplines. Assessment development is also a key part of the modules and we are building on what we have discovered using the ECD process as described herein. The work to date is also making clear how much more there is to do to identify, assess, and instruct multiple source comprehension and better prepare students for the complex literacy demands of the 21st century.

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. Thanks to James Pellegrino for his guidance on assessment design. The assessment project described in this chapter was funded, in part, by the Institute for Education Sciences, U.S. Department of Education (Grant R305G050091). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring organization.

Notes

1 Preliminary analyses indicated that none of these trends was significantly related to grade; nor was membership in the clusters.
2 Project READI (Reading, Evidence, and Argumentation in Disciplinary Instruction) is a collaborative project among researchers from the University of Illinois at Chicago, Northern Illinois University, Northwestern University, WestEd, and Inquirium Ltd. who are working with educational practitioners from schools in the greater Chicago and Bay area in California. Project READI is part of the Reading for Understanding network funded by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100007 to the University of Illinois at Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

students’ use of source attributes to select useful sources. *Journal of Educational Computing Research, 41*, 63–82.


In B. K. Britton & A. C. Graesser (Eds.), *Models of text comprehension* (pp. 73–113). Hillsdale, NJ: Erlbaum.


