Running head: READING INSTRUCTION

Technologies that Support Reading Instruction

Arthur C. Graesser and Danielle S. McNamara
University of Memphis

Graesser, A.C., & McNamara, D.S. (in press). Technologies that support reading comprehension. In C. Dede and J. Richards (Eds.), *Digital teaching platforms*.

Send correspondence to:

Art Graesser
Psychology Department
202 Psychology Building
University of Memphis
Memphis, TN, 38152-3230
901-678-4857
901-678-2579 (fax)
a-graesser@memphis.edu

KEYWORDS : Discourse processing; comprehension,

Technologies that Support Reading Comprehension

The purpose of this chapter is to provide a snapshot of technologies that support individualized reading instruction. We concentrate on the technologies that are currently available and within the future horizon rather than reconstructing the technologies of the past. Teachers will presumably be expected to incorporate these technologies in classrooms and small group projects in addition to one-on-one tutoring of individual students. This chapter also emphasizes technologies that help students comprehend text beyond the word and into the realm of sentence and discourse meaning. This is not to deny the importance of word decoding and vocabulary, both of which are addressed in this chapter. However, the deeper levels of language and discourse that have been particularly mysterious in the arenas of research, teaching, evaluation, and policy. It is the deeper levels that have been the primary focus of our own research and technology development.

The reading experience 50 years ago was less complex than it is today in our world of multimedia, multitasking, and multi-communication. Once upon a time children and adults had the opportunity to read books in a linear fashion, beginning to end, in a single sitting, over long stretches of time. This opportunity is a rare luxury in today's world. Nevertheless, all students in today's *No Child Left Behind* and *Race to the Top* are expected to read proficiently, including at deeper levels of comprehension. It is well documented that reading literacy opens the doors to better professions, higher salaries, and an improved quality of life (Resnick, 2010). But how can that be accomplished in today's world? Today we live in a world of email, instant messaging, facebook, chat rooms, portals, Google, Wikipedia, teleconferences, solitary and multiparty games, sensuously rich video, UTube, Twitter, iPhones, and other technologies that break up our experiences into smaller packages of time and content. Texts are not only shorter, but the reading process is typically distributed and integrated with other media, tasks, and actions. Texts are read for the purpose of accomplishing short-term goals, solving problems, communicating with others, writing reports, and playing games. Given today's information ecology, it is becoming more challenging to place students in an environment where they can have long periods of sustained concentration, which possibly is a prerequisite for acquiring the skills of reading at deeper levels of comprehension. But perhaps it is not a prerequisite. Perhaps deep comprehension can be achieved by a distributed information ecology that is integrated with action and guided by reader goals.

This chapter is divided into four sections. The first section identifies the many levels of language and discourse that need to be considered in a comprehensive model of reading intervention. A reader may have deficits at one or more of these levels so the technology needs to diagnose particular deficits for particular readers. The second section presents some automated metrics for scaling texts on *text complexity* at multiple levels of language and meaning. The assignment of texts to readers can be made in a fashion that fits the readers' proficiency profiles in addition to their self-regulated choice of materials to read. The third section presents some examples of technology that can diagnose reading deficits and train students to rectify such deficits. These technologies typically follow a *diagnosis and remediation* framework that is tailored to individual readers. The fourth section explores the reading process in more advanced digital environments that consider student motivation and that incorporate multiple media, tasks, and pedagogical objectives. In this case, students read in order to learn, act, solve problems, and communicate with others. One prevailing question that surfaces throughout this chapter is how teachers and other professionals can use these new technologies effectively in a digital teaching platform.

Levels of Language, Discourse, and Reading Comprehension

Psychological theories of comprehension have identified the representations, structures, strategies, and processes at multiple levels of language and discourse (Graesser & McNamara, 2010; Kintsch, 1998; McNamara & Magliano, 2009). This chapter considers six levels, as elaborated in Table 1: *words*, *syntax*, the explicit *textbase*, the referential *situation model* (sometimes called the mental model), the discourse *genre and rhetorical structure* (the type of discourse and its composition), and the *pragmatic communication* level (between speaker and listener, or writer and reader). Words, syntax, and the pragmatic communication need not be elaborated here because they are self explanatory, whereas the other levels of meaning are a bit more esoteric. The textbase contains explicit ideas in the text in a form that preserves the meaning but not the precise wording and syntax. The situation model is the subject matter that is being described in informational texts or the microworld that evolves in a story. This would include the people, objects, spatial setting, actions, events, processes, plans, thoughts and emotions of people, and other referential content. The text genre is the type of discourse, such as a news story, a folk tale, a persuasive editorial, or a science text that explains a causal mechanism. The rhetorical structure is the organization of the text at a macro-level and the discourse function of particular excerpts. Table 1 elaborates on all six levels by identifying the codes, constituents, and content associated with each level. This chapter does not crisply define each level and the associated terminology, but Table 1 does give example components of what each level contains.

Table 1 depicts the levels as compositional components that are constructed as a *result* of successful comprehension. This *compositional* viewpoint alone is insufficient for researchers who are interested in deficits in comprehensions and interventions to remediate such deficits. Each component C also has the following viewpoints:

Knowledge. For any given component C, the reader needs to have the prerequisite knowledge about C through prior experience and/or training.

Process. The reader needs to be able to process C by recognizing patterns in the text and by executing relevant skills and strategies proficiently.

Diagnosis of deficit. If the reader is not proficient in processing component C, the technology or teacher need to diagnose the deficit.

Remediation of deficit. The technology or teacher attempt to remediate the deficit by some form of training or intervention.

Compensation of deficit. Some readers show no response to an intervention that attempts to remediate the deficit. In that case, there is a need to compensate for the deficit through alternative reading components or augmented technologies.

INSERT TABLE 1 ABOUT HERE

Readers can face reading obstacles at any of the levels in Table 1. There can be deficits in the reader (e.g., lack of knowledge or skill), the text (e.g., incoherent text, esoteric irrelevant jargon), or the training (e.g., an emphasis on shallow levels of reading). The severity of an obstacle can range from a minor irregularity that adds some cost in processing time to a major impasse that results in a complete breakdown in comprehension. Attempts can be made to compensate for the problem by recruiting information from other levels of discourse, from prior knowledge, from external sources (e.g., other people or technologies), or from strategies. The scenarios below illustrate some obstacles and resulting consequences.

Scenario 1. A child has trouble recognizing letters in the alphabet so there is an obstacle in lexical decoding at the word level (level 1 in Table 1). The word deficit blocks him from understanding any of the text at levels 2-6.

Scenario 2. A high school student reads a health insurance document that has lengthy sentences with embedded clauses, numerous quantifiers (*all, many, rarely*), and many logical operators (*and, or, not, if*). She understands nearly all of the words, but has only a vague idea what the document explicitly states because of complex syntax, a dense textbase, and an ungrounded situation model (i.e., deficits at levels 2-4). However, she signs the contract because she understands its purpose and trusts the school. Levels 5 and 6 circumvent the need to understand levels 2-4 completely.

Scenario 3. Laboratory partners in an engineering course read the directions to assemble a new computer. They argue about how to hook up the cables on the dual monitors. They have no problem understanding the words and textbase in the directions (levels 1-3) and no problem understanding the genre and purpose of the document (levels 5 and 6), but they do have a deficit at the situation model level (level 4).

Scenario 4. A science student asks his roommate to proofread a term paper, but the roommate is a journalism major who knows little about science and complains that there is a problem with logical flow. The science major revises the text by adding connectives (e.g., *because, so, therefore, before*) and other words to improve the cohesion. The revised composition is deemed more comprehensible. In this case, improvements at levels 1 and 3 compensated for a deficit at level 4.

. **Scenario 5**. Parents take their children to a new Disney movie that they discover has a few adult themes. The children notice the parents laughing at different points in the movie than they do. The children are making it successfully through discourse levels 1-4, but levels 5 and 6 are not intact.

These scenarios illustrate how deficiencies at one or more discourse levels can have substantial repercussions on the processing at other levels. Reading researchers need to understand the processing mechanisms both *within* levels and *between* levels.

Obstacles to comprehension are often invisible to the reader. Research on meta-cognition has extensively documented that most adults have limited abilities to detect and monitor many cognitive states (Hacker, Dunlosky, & Graesser, 2009). Research on *comprehension calibration* has collected ratings from readers on how well they believe they have comprehended texts and these ratings are correlated with objective tests of text comprehension. The comprehension calibration correlations are alarming low (r = .27) even among college students (Maki, 1998). Readers often have an illusion of comprehension when they read text because they settle for shallow levels of analysis as a criterion for adequate comprehension (Daneman, Lennertz, & Hannon, 2007). Shallow readers believe they have adequately comprehended text if they can recognize the content words and can understand most of the sentences, when in fact they are missing the deeper knowledge, occasional contradictions, and false claims. Deep comprehension requires inferences, linking ideas coherently, scrutinizing the validity of claims with a critical stance, comparing content of different texts, and understanding the motives of authors (Kendeou & Van den Broek, 2007; Rouet, 2006; Wiley et al., 2009). Deep comprehension may only be selectively achieved in everyday reading experiences.

In studies conducted in our laboratory (Graesser et al., 2004; VanLehn et al., 2007), college students read textbooks on technical topics such as computer literacy and Newtonian physics. They subsequently complete a test on deep knowledge with multiple-choice questions similar to the Force Concept Inventory in physics (Hestenes, Wells, & Swackhamer, 1992) or through essays that require deep reasoning. The results revealed that students had zero learning gains from reading the textbook and that the posttest scores did not differ from a condition in which the students read nothing at all. In contrast, the learning gains were quite substantial in tests of shallow knowledge and also in tests of deep knowledge when there was a learning environment that challenged their comprehension of the material (through a computer system called AutoTutor, an animated agent that holds conversations with the reader in natural language). Results such as these support the general claim the reading strategies of literate adults are often aligned with shallow rather than deep comprehension even when they have a nontrivial amount of world knowledge on the topics and when they have sufficient reading strategies to land them in college. Readers need to be pushed into deeper waters through challenges from the text, tests, technology, or teacher.

From the standpoint of this volume, a fundamental question is how teachers can manage technologies to help students improve their reading and achieve deeper levels of comprehension. Quite clearly, the teachers will need to understand the many manifestations of meaning. The six levels in our multilevel framework is only a first step that needs to be fleshed out and expanded. Aside from this professional development on the components of meaning, the technology needs to inform the teacher what problems are experienced by particular readers and how to remediate the deficits. Technologies are available to meet these challenges, but teachers need to understand how to use them effectively.

Computer Tools for Analyzing Text at Multiple Levels

One way to challenge students is to assign them texts that push the envelope on what they can handle. The texts presumably should not be too difficult or too easy, but rather at the right zone of complexity. The texts should gracefully push the frontier of their personal proficiency profiles -- at their zone of proximal development. In order to do this, we need to be able to scale the texts on complexity at multiple levels of comprehension.

This is a unique point in history because there is widespread access to computer tools that analyze specific texts and large text repositories (called *corpora*). This increase in automated text analyses can be attributed to landmark advances in computational linguistics (Jurafsky & Martin, 2008), statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007), and corpus analyses (Biber, Conrad, & Reppen, 1998). Thousands of texts can be quickly accessed and analyzed on thousands of measures in a short amount of time.

One approach to scaling texts is to have a single dimension of text complexity. This is the approach taken by metrics such as Flesch-Kincaid Grade Level (Klare, 1974-5), Degrees of Reading Power (DRP; Koslin, Zeno, & Koslin, 1987), and Lexile scores (Stenner, 2006). The three metrics of text complexity are highly correlated ($r > .90$). The Flesch-Kincaid Grade Level metric is based on the length of words and length of sentences. DRP and Lexile scores relate characteristics of the texts to performance of readers in a cloze task. In the cloze task, the reader first reads the text and then is presented with the text with words left blank; the reader is asked to fill in the words by generating them or by selecting a word from a set of options. A text is at the reader's level of proficiency if the reader can perform the cloze task at a threshold of performance (75%). A text is defined as easy for a population of readers if performance exceeds 75% and difficult to the extent it is lower than 75%.

The overall metric of text difficulty is limited in its utility in helping students improve their reading, however. There needs to be more specific guidance on what is wrong and how to fix it. To fill this gap, we have developed an automated text analysis system called Coh-Metrix

(Graesser & McNamara, 2010; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, McCarthy, & Graesser, in press). Coh-Metrix is a computer facility that analyzes texts on most of the discourse levels in Table 1. More specifically, Coh-Metrix was developed to analyze and measure text on levels 1 through 5. Our goals were: (a) to enhance standard text difficulty measures by providing scores on various cohesion and language characteristics and (b) to determine the appropriateness of a text for a reader with a particular profile of cognitive characteristics. Coh-Metrix is available in both a public version for free on the web (http://cohmetrix.memphis.edu, version 2.0) and an internal version (versions 2.1 and 3.0). The public version has over 60 measures of language and discourse at levels 1-5 in Table 1, whereas the internal research version has nearly a thousand measures that are at various stages of testing.

It is beyond the scope of this chapter to list and define the measures of Coh-Metrix. Table 2 lists many of them whereas definitions and elaborations of these measures are in the web site and publications listed above. In an ideal world, students could be assessed on their proficiency on several dozen reading components and the remediation modules could be pitched at a fine grain. However, it would be impractical to expect teachers, students, and school administrators to monitor such detail. There needs to be an ideal sweet spot between a single dimension on text complexity and dozens/hundreds of specific metrics by Coh-Metrix.

INSERT TABLE 2 ABOUT HERE

A recent Coh-Metrix analysis was performed on 37,651 texts in a corpus provided by TASA (Touchstone Applied Science Associates). This TASA corpus represents the texts that a typical senior in high school would have encountered throughout schooling in K12. That is, there are texts that represent kindergarten through 12th grade. These texts are scaled on Degrees of Reading Power, which can approximately be translated into grade level. The vast majority of the texts are between 250 and 350 words in length. Most of the text genres were classified by TASA as being in language arts (narrative), science, and social studies/history, but other categories were business, health, home economics, and industrial arts.

We have performed statistical analyses (called principal components analysis) on the TASA text corpus in order to discover what aspects of texts account for text complexity. We explored over 100 measures of Coh-Metrix and included 53 in the final analysis. We discovered that 8 dimensions could account for 67% of the variability among texts, which is extremely impressive. Five of these dimensions accounted for most of the results. These five dimensions are listed and defined below.

1. **Narrativity.** Narrative text tells a story, with characters, events, places, and things that are familiar to the reader. Narrative is closely affiliated with everyday oral conversation.
2. **Referential cohesion**. High cohesion texts contain words and ideas that overlap across sentences and the entire text, forming threads that connect the explicit text together for the reader.
3. **Situation model cohesion.** Causal, intentional, and temporal connectives help the reader to form a more coherent and deeper understanding of the text.
4. **Syntactic simplicity**. Sentences with few words and simple, familiar syntactic structures are easier to process and understand. Complex sentences have structurally embedded syntax.
5. **Word concreteness**. Concrete words evoke mental images and are more meaningful to the reader than abstract words.

Each of the 5 dimensions are above are expressed in terms of ease of comprehension. Text complexity is defined as the opposite of ease, so principal component scores are reversed in measures of text complexity.

For each text, we were able to compute a score on each of the 5 principal components. These were measured as z-scores, a standardized metric in standard deviation units. A z-score of 0 is average for all of the 37,651 texts. Scores are higher and more positive when the texts are easier on the component and more negative when the texts are more difficult. Figure 1 presents mean z-scores as a function of the DRP-based grade levels (1-12). There is a line for each of the 5 principal components.

The z-scores in Figure 1 support a number of conclusions. First, there are at last 5 dimensions of text complexity that should be considered when analyzing texts. These dimensions have a foundation in theories of reading comprehension that were described earlier (Graesser & McNamara, 2010; Kintsch, 1998) in addition to explaining a large corpus of texts. Indeed, the close alignment between theory and the inductive principal component analysis was an unexpected, if not uncanny, confirmation of the validity and value of our multilevel theoretical framework. Second, the DRP grade levels are most correlated with narrativity and syntax. Texts at higher grade levels tend not to be narrative texts, but rather are more technical informational texts, and tend to have more complex syntax. Second, the two cohesion measures tend to slightly increase as a function of DRP grade level, but the correlation is more modest. Standard text complexity measures do not highly correlate with these cohesion measures, which suggests that unidimensional measures are not tapping more global discourse meaning. Indeed, nearly all standard text complexity metrics handle word and sentence constraints, but not deeper and more global meaning. Interestingly, cohesion is comparatively low at early grade levels, so there is a worry that early readers suffer from low cohesion (McNamara et al., in press). And third, word concreteness tends to stay constant until high school, where there is a higher amount of abstract words.

There is the practical question of how these complex analyses can be communicated to teachers, students, and administrators who may not be highly trained in research methods. One approach is to convert the z-scores to percentile scores on text complexity, which are likely to be more familiar to teachers and citizens. A percentile score varies from 0 to 100%, with higher scores meaning the text is more complex. For example, a percentile score of 80% means that 20% of the texts are more difficult and 80% less difficult.

Text complexity profiles are presented in Figure 2 for two example science texts on physics. We see that the *Gravity Reverse* passage is a much higher grade level than the *Elementary Particles* passage. However, difficulty scores on the 5 components show a different picture. The Elementary Particles passage is actually somewhat more difficult when one considers the deeper components of text genre (more informational than narrative) and situation model cohesion. Profiles such as these can be viewed by teachers when they suggest texts for students to read. They can be viewed by students when they choose what to read in digital libraries or in game environments that assign more points for texts that are more challenging. The profiles can be viewed by principals and administrators when they select textbooks and other readings for students in different achievement clusters.

The text complexity profiles can also be generated automatically by the computer for any text. This opens the door for an automated selection of texts to assign to readers, to suggest that the readers read, or to include in a library for the readers to select on their own. The automated text selection mechanism would be sensitive to the reader’s profile of reading proficiency at various levels. The texts would push the envelope at the reader's challenge zone for the various components, based on what the computer knows about the reader and about the text. The computer algorithm can also consider motivational, emotional, and non-cognitive characteristics of the reader. Self-efficacy and academic risk taking are particularly relevant non-cognitive variables (Meyer & Turner, 2006). Readers with high self-efficacy (i.e., they are convinced they can perform well) and academic risk (i.e., they take on challenging tasks and don't get emotionally upset if they fail) can be encouraged to read texts that aggressively expand their challenge zone. However, readers with low self-efficacy and academic risk taking would be given texts at the edge or within the challenge zone to ensure they succeed.

INSERT FIGURES 1 AND 2 ABOUT HERE

Once again, teachers would need to be trained how to use an automated text analysis and assignment system. They would need to understand what the reading components mean, what the graphical data depict, and how the texts are assigned if the teachers are deeply involved in the process. Some of these mechanisms will and should remain invisible to the teacher and student, however. A central question for professional development of teachers is to identify features of the text analysis and assignment systems that should be visible to the teacher and should inform the teacher's decisions.

Interventions for Reading Instruction

For decades, a *diagnosis and remediation* framework has guided intervention programs for reading instruction. The first step is to assess how well the student has mastered particular skills, strategies, and knowledge relevant to reading. These reading proficiencies are assessed by psychometric tests that are administered by professional testers or by computer. The assessments have traditionally addressed proficiencies that funnel into one of three categories: lexical decoding, vocabulary, and meaning. Lexical decoding includes such generic skills as the identification of alphabetic symbols, the association between speech sounds and letters, phonemic awareness, decomposition of words into syllables or morphemes, and speed (i.e., fluency) of accomplishing these subskills (Adams, 1990). There are generic tests of vocabulary, which measure how many words a student knows. However, sometimes particular texts have rare words and esoteric jargon that require pre-training before the student has a fair chance at comprehending a text. Available research has also confirmed that a word is not mastered by simply knowing its definition; a word needs to be experienced in multiple contexts of use before it is conceptually rich enough to guide the interpretation of sentence meanings and the generation of inferences (Perfetti, 2007). Regarding meaning, tests such as the Woodcock-Johnson and Gates-MacGinitie Reading Tests, are routinely administered to readers in order to assess meaning comprehension (in addition to other components of reading and cognition). However, the analysis of meaning is not as fine-grained and theoretically grounded as we are proposing in this chapter.

There is ample evidence that individualized student instruction (ISI) is more effective than having all students in a classroom move at the same pace on the same materials (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007). The reason is because there are aptitude-treatment interactions: some types of readers benefit from one instruction method whereas other types of readers benefit from another. One powerful demonstration of this is reported by Connor et al. (2007, 2009) who developed and tested her *Assessment to Instruction* (A2i) web-based software. She compared students who were above versus below average in lexical decoding (i.e., letter and word reading skills) and students with above versus below average in vocabulary. She discovered that the following instruction methods were appropriate for different types of students.

IF (low lexical decoding) → THEN (use explicit teacher-managed code-focused instruction)

IF (high lexical decoding) → THEN (don't use explicit teacher-managed code-focused instruction)

IF (low vocabulary) → THEN (use combination of explicit teacher-managed code-focused instruction and explicit meaning-focused instruction)

IF (high vocabulary) → THEN (use explicit meaning-focused instruction or independent reading)

Students with high lexical decoding skills and vocabulary may best be left alone to conduct independent reading. Reading on topics that interest them would no doubt be a high priority, but sometimes in practice the teacher would be expected to nudge them to topics that address state standards in science and other areas. In contrast, more explicit instruction is needed for the other groups of readers.

There is an abundance of computer technologies available for instruction at the level of lexical decoding and vocabulary. For example, text-to-speech engines can produce spoken output as the words being spoken are highlighted on the computer display. On the flip side, in speech-to-text facilities the student reads text aloud and the system gives feedback on their errors (Mostow, 2008). In multimedia environments there are simultaneous presentations of printed words and their pictorial referents in the form of static images (e.g., a picture of a truck) or dynamic events (shooting a basketball). There is a healthy cottage industry of computer environments to train students to associate sounds, letters, words, and pictures, to decompose complex words into its syllables or morphemes, and to help students learn vocabulary. The bigger challenge has been to develop computer environments with instruction at the level of meaning. Nevertheless, these learning environments have begun to emerge in the last decade (see edited volume by McNamara, 2007).

One promising system that focuses on deeper levels of meaning is the iSTART system (McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). iSTART is a web-based computer program that uses conversational agents to provide reading strategy training for deeper levels of comprehension. iSTART incorporates theoretically motivated *Self-Explanation Reading Training* (SERT, McNamara, 2004) which has proven to be effective when used by teachers in classrooms. SERT and iSTART teach students to self-explain science texts by using active reading strategies that facilitate and enhance comprehension. These strategies include paraphrasing explicit text, generating elaborative inferences, generating bridging inferences that connect text elements, predicting what will happen next in the text, and monitoring one's comprehension.

The original version of iSTART had three modules geared toward teaching students to use self-explanation and reading strategies while reading challenging texts. In an *Introduction* module, students watch the teacher-agent explain the reading strategies to two student-agents. In a *Demonstration* module, students are quizzed on various aspects of the strategies as they communicate with interactions with agents. In a *Practice* module, students practice generating typed self-explanations and the conversational agents provide feedback on performance. The practice section incorporates feedback that to some extent is adaptive to the level of student performance. For example, a paraphrase strategy is appropriate when the student has low ability and the self-explanations are not remotely related to the text. In contrast, bridging and elaborative inference strategies would be better suited to students with greater world knowledge and reading ability. Numerous experiments assessing the efficacy of iSTART have been conducted with over 1,000 middle school, high school, and college students. iSTART is effective in helping students use strategies to learn from texts. It also improves comprehension, particularly among low-knowledge readers (McNamara, O'Reilly, Best, & Ozuru, 2006; O'Reilly & McNamara, 2007).

More recent versions of iSTART have added features to further improve learning and motivation. The strategy training that is more adaptable to students' needs and rates of progress. The system has more intelligent mechanisms to track the reader's ability and to respond with appropriate instruction. There is a game version, called iSTART-ME (Motivationally Enhanced), that has laces in game elements to give the students an enhanced sense of control, performance-based feedback and reinforcement, challenges, and other features that optimize motivation.

One major challenge is to accommodate the teacher's course demands. In essence, there needs to be a computer-human software interface that allows teachers to integrate iSTART into their classrooms. Incorporating a computerized tutor into a classroom is not as simple as merely giving the system to the teacher and expecting that it will be used consistently or successfully. There are many constraints that must be met in order for the system to be integrated into classroom in an effective manner. First, the teacher must understand the need for reading strategy training and to be receptive to intelligent tutoring systems. Second, the program must be easy for the teacher to use, with facilities that handle the teachers' questions about the program and data on the students' progress. A subcomponent called the teacher interface was developed to facilitate teachers' use of iSTART in the classroom. Third, the number of course topics and range of text difficulty covered during the practice sessions needed to be increased to make the system applicable to a wide variety of educational topics and students with varying levels of ability.

The above considerations motivated the iSTART team to develop four parts of the interface for the teachers.

(1) *Teacher Interface Front Page*. This is a front page to the modules that are available to teacher.

(2) *Teacher Introduction Module (TIM).* This section provides brief discussions and overviews of cognition, text comprehension, the iSTART trainer, and iSTART reading strategies. It is recommended that the teacher read through the topics in order to understand the theory behind iSTART.

(3) *Training Organizer and Manager Module (TOM).* The goal of this module is to provide technical and practical support to the teachers so that they can use iSTART effectively. The program has default sequences for students based on their assessed ability; however, this module allows teachers to make special assignments and constrain iSTART's use of its library to best serve their classes' needs. We have built a library of texts that encompasses a wide selection of topics and domains from which to choose and have built tools that make it easy to add additional texts. The curriculum policy concerns whether the students take the modules at their own pace, one after the other, or take them on a schedule.

(4) *Performance Analyzing Tool Module (PAT).* The purpose of this module is to provide teachers with a means to monitor students' performance and progress during training. This tool reports the students' pre-training assessment scores, assessments that occur

during the iSTART training, and the students' progress through the iSTART curriculum. These reports help teachers gauge the students' performance and recommend remediation interventions.

Teachers are not currently trained to handle all of these levels of complexity. The major challenges will be to develop adequate authoring tools and to improve professional development to fill these gaps. The field is currently in flux on how to solve these challenges.

Reading for Learning and Doing

As a student progresses through K12 there is a shift from learning-to-read to reading-to-learn. Reading becomes more of a goal-driven activity (McCrudden & Schraw, 2007) that supports other activities. Sometimes the reader has passionate interests and intrinsic goals, such as a hobby, a pet project, or a multi-party game with a community of peers. More often there are projects directed by the teacher, such as a writing assignment, a science exhibit, or the design of an artifact. In both cases, the student needs to hunt for reading materials that are relevant to the goals. When an article is accessed, the reader assesses the relevance of sections read, searches for answers to specific questions, and exits the text when it fails to deliver. The half-life of reading an article is very short for all but a small minority of texts that are particularly relevant, insightful, and helpful.

Researchers need to collect more data on the patterns of reading activities of readers throughout the day in both school and informal settings. An ecological profile of reading behavior would be tremendously insightful. We do have some knowledge of how students read digital libraries and the Internet under limited contexts with goals forced upon the reader. For example, Wiley et al. (2010) collected eye tracking data and/or think aloud protocols as college students examined a set of web sites on plate tectonics for the purpose of writing an essay on what caused the eruption of Mt. St. Helens volcano. The deeper comprehenders of plate tectonics did spend more time on high quality web sites with rigorous science than on web sites with pseudoscience (e.g., earthquakes are cause by oil drilling, or the alignment of planets in the solar system). They also spent more time on information on pages that are relevant to the goal and focused more on explanations of the scientific mechanism. One wonders, however, whether this laboratory reading profile matches their reading habits in the wild.

It is difficult for most readers to build *causal* mental models of topics in science and history, even after reading multiple texts (Rouet, 2006; Wiley et al., in press). Sometimes there are simple linear causal chains of events, but more often there are complex enabling states, tradeoffs between causes, cycles of events, and dynamical systems that add considerable complexity (Grotzer, 2003). Students are prone to avoid such complex material unless there is a tremendous push or incentive in the learning environment. Serious multiparty games provide one context for encouraging such deeper learning.

Dede and Grotzer (2009) have recently developed an ecoMUVE curriculum that incorporates the National Science Education Standards for life sciences in grades 6 and 7. MUVEs are similar to some online multi-player games in that they enable multiple participants to access virtual worlds simultaneously and to utilize digital artifacts. Participants navigate these worlds through their graphical representations (avatars) and interact with other students and with computer-based agents to facilitate collaborative learning activities of various types. Due to their immersive nature, MUVEs are an effective platform for providing authentic inquiry in middle school science. EcoMUVE comprises two one-week modules to complement and extend the current 4-week curriculum for ecosystems in grades 6-7. Students explore the ecosystem of a pond and surrounding area to see realistic organisms in their natural habitats. A submarine tool allows students to collect the microscopic organisms in the pond, helping students understand that ecosystems also involve non-obvious causes that are hard to detect with the naked eye. A food web tool allows students to build a food web of the ecosystem populations. Students interact with the domino causality of food web relationships, helping them move beyond noticing only direct effects. Students visit the virtual pond over time, with the opportunity to observe, explore, and collect physical and chemical data. On one simulated day, students make the surprising discovery of a fish kill. Students are challenged to figure out what happened by collecting and analyzing information to solve the mystery and to understand the complex causality of the pond ecosystem. An analysis of the ecosystem poses many challenges about causality, such as distal causes, tradeoffs between variables, relationships between micro and macro levels of explanation, counterintuitive consequences of events, and dismissal of folklore explanations. EcoMUVE encourages deeper learning about causal mental models by challenging them at deeper conceptual level, by placing them in *cognitive disequilibrium* or *desirable difficulties* as they face these challenges (Bjork & Linn, 2006; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Schwartz & Bransford, 1998), and by requiring them to construct explanations, arguments, justifications, clarifications, and so on as they communicate with peers on relevant subject matter content.

From the standpoint of this chapter, there is a host of questions about the students reading activities as they experience ecoMUVE. There is a digital library of articles, most of which are relevant to the pond ecosystem. What goals do students pursue and do they read articles that are relevant to their goals? When an article is fetched, what do they read, how much do they read, and when do they give up? How much of the information in an article gets incorporated in messages to peers, documents they write, arguments, and behavior? What deficits in reading components (see Table 1) present barriers to effective participation in ecoMUVE? Answers to all of these questions can be explored by automatic tracking of the behavior and communication of the students who interact with ecoMUVE. These data are available in log files and can be submitted to data mining and data farming analyses.

Quite clearly, the role of the teacher in ecoMUVE and other advanced learning environments will require some radical changes in professional development. Mentors are needed with expertise in pedagogy in addition to those with expert subject matter knowledge if any deep learning is to occur (Shaffer, 2006). The teachers who manage these educational technologies will have a radically different skill set than the teachers of 50 years ago. It is easy to assign a book, prepare a test, and grade it. It is difficult to understand and use the technologies discussed in this chapter.

Acknowledgements

The research on was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918, ALT-0834847, DRK-12-0918409), the Institute of Education Sciences (R305G020018, R305H050169, R305B070349, R305A080589, R305A080594), and the Department of Defense Counter Intelligence Field Activity (H9C104-07-0014). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, or DoD.
Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230, a-graesser@memphis.edu.

References

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: Introducing desirable difficulties. *American Psychological Society Observer*, *19*, 29,39.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: Algorithm-guided individualized reading instruction. *Science, 315*, 464-465.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development, 80*, 77-100.

Daneman, M., Lennertz, T. &Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes,22*,83-105.

Dede, C., & Grotzer, T. (2009). *Black's Nook EcoMUVE: Learning about ecosystems complexity: Teacher guide*. Technical Report on Institute for Education Sciences Grant R305A080514, Harvard University.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers,* 36, 180-193.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, *33*, 1235–1247.

Graesser, A.C., & McNamara, D.S. (2010). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193-202.

Grotzer, T. A. (2003). Learning to understand the forms of causality implicit in scientific explanations. *Studies in Science Education, 39*, 1–74.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition in education*. Mahwah, NJ: Erlbaum/Taylor & Francis.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141-158.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Englewood, NJ: Prentice Hall.

Kendeou, P. & van den Broek, P. (2007). The effects of prior knowledge in text structure on comprehension processes during reading scientific texts. *Memory & Cognition, 35*, 1567-1577.

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

Klare, G. R. (1974–1975). Assessing readability. *Reading Research Quarterly, 10*, 62-102.

Koslin, B.I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York: College Entrance Examination Board.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007)(Eds.). *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Erlbaum.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.

McCrudden, M.T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*, 113-139.

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

McNamara, D. S. (2007)(Ed.) *Reading comprehension strategies: Theories, interventions, and technologies*. Mahwah, NJ: Erlbaum.

McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397–421). Mahwah, NJ: Erlbaum.

McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (in press). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*.

McNamara, D.S., & Magliano, J. (2008). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 51.

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*, 147-171.

Meyer, D. K., & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, *18,* 377-390.

Mostow, J. (2008). Experience from a Reading Tutor that listens: Evaluation purposes, excuses, and methods. In C. K. Kinzer, L. Verhoeven (Eds.), *Interactive Literacy Education: Facilitating Literacy Environments Through Technology (pp. 117-148).* Mahwah, NJ: Erlbaum

O'Reilly,T. & McNamara, D.S.(2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement.*American Educational Research Journal*, *44*,161-197.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading,11*, 357-383.

Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher.*

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

Schwartz, D.L., & Bransford, J.D. (1998). A time for telling. *Cognition & Instruction, 16(4),* 475-522.

Shaffer, D. W. (2006). *How Computer Games Help Children Learn*. Palgrave Macmillan

Snow, C. B. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND.

Stenner, A.J.(2006). *Measuring reading comprehension with the Lexile framework*. Durham, NC: Metametrics, Inc. presented at the California Comparability Symposium, October 1996. Retrieved January 30, 2006 from http://www.lexile.com/DesktopDefault.aspx?view=re.

VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., & Rose, C.P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*, 3-62.

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

Table 1: *Levels of Language and Discourse*

LEVEL	EXAMPLE COMPONENTS OF LEVEL
Words	Word meaning representation Word composition (graphemes, phonemes, syllables, morphemes, lemmas) Parts of speech (noun, verb, adjective, adverb, determiner, connective)
Syntax	Syntax (noun-phrase, verb-phrase, prepositional phrase, clause) Linguistic style
Textbase	Semantic meaning Explicit propositions or clauses Referring expressions linked to other text constituents Bridging inferences that connect explicit propositions, clauses, or words
Situation model	Situation conveyed in the text Agents, objects, and abstract entities Dimensions of temporality, spatiality, causality, intentionality Inferences that elaborate text and link to the reader's experiential knowledge Connectives that explicitly link events, actions, states, and goals Given versus new information Images and mental simulations of events
Genre & rhetorical structure	Discourse category (narrative, persuasive, expository, descriptive) Rhetorical composition (cause+effect, claim+evidence, problem+solution) Epistemological status of propositions and clauses (claim, evidence, warrant) Speech act categories (assertion, question, command, request, greeting, etc.) Theme, moral, or point of discourse
Pragmatic communication	Goals of author Attitudes and beliefs (humor, sarcasm, eulogy, deprecation)

Table 2: *Example Coh-Metrix Measures and Indices (over 700 available)*

LEVEL OR CLASS	MEASURE (INDEX)
Words	Frequency, concreteness, imagery, age of acquisition, part of speech, content words, pronouns, negations, connectives (different categories), logical operators, polysemy, hypernym/hyponym (reflects abstractness); these counts per 1000 words.
Syntax	Syntactic complexity (words per noun-phrase, words before main verb of main clause)
Textbase cohesion	Cohesion of adjacent sentences as measured by overlapping nouns, pronouns, meaning stems (lemma, morpheme). Proportion of content words that overlap. Cohesion of all pairs of sentences in a paragraph.
Situation model cohesion	Cohesion of adjacent sentences with respect to causality, intentionality, temporality, spatiality, and latent semantic analysis (LSA). Cohesion among all sentences in paragraph and between paragraphs via LSA. Given versus new content.
Genre and rhetoric	Type of genre (narrative, science, other). Topic sentencehood
Other	Flesch-Kincaid grade level, type token ration, syllables per word, words per sentence, sentences and paragraphs per 1000 words.

Figure 1: Z-scores on Five Coh-Metrix Components as a Function of DRP Grade Levels

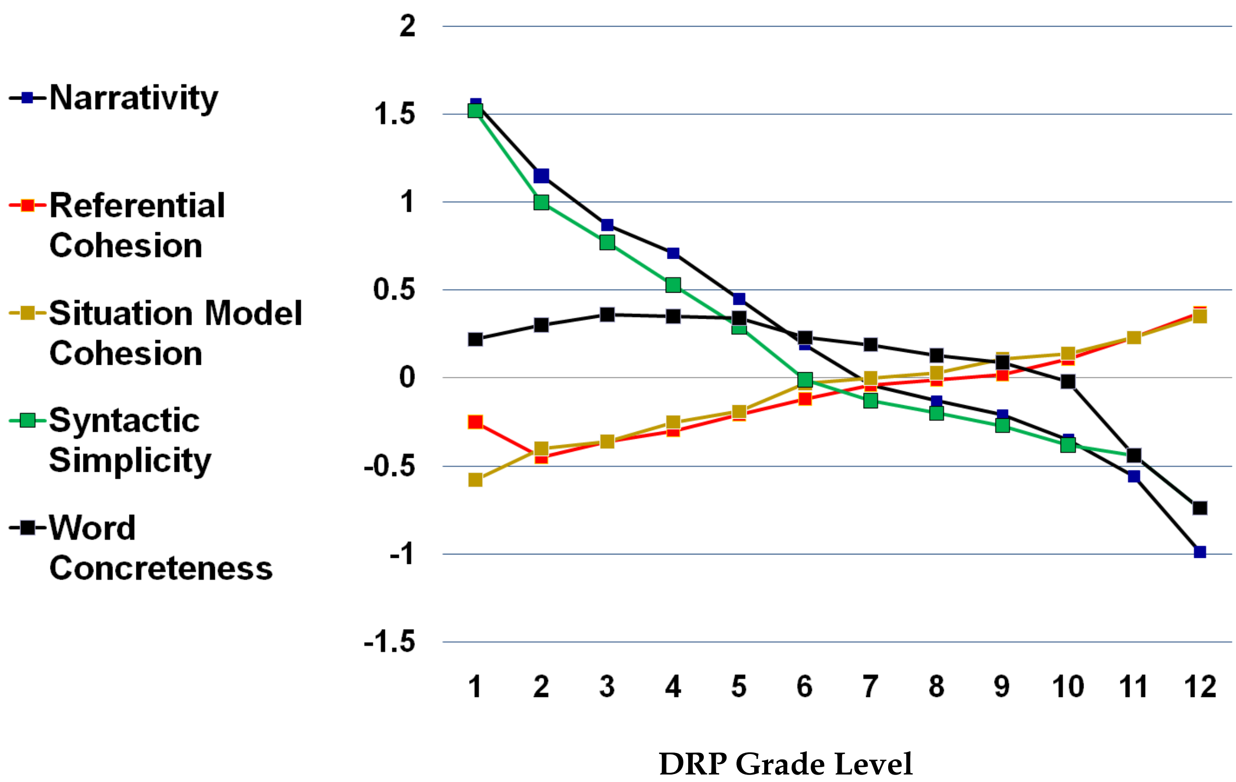

Figure 2: Coh-Metrix Text Complexity Profiles for Two Science Texts.