

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language, 31*, 634–647. doi:10.1016/0749-596X(92)90032-S

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. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46*, 1060–1106. doi: 10.3102/0002831209333183

Learning, Thinking, and Emoting With Discourse Technologies

Arthur C. Graesser
University of Memphis

This is an unusual moment in the history of psychology because of landmark advances in digital information technologies, computational linguistics, and other fields that use the computer to analyze language, discourse, and behavior. The technologies developed from this interdisciplinary fusion are helping students learn and think in ways that are sensitive to their cognitive and emotional states. Recent projects have developed computer technologies that help us understand the nature of conversational discourse and text comprehension in addition to improving learning. AutoTutor and other systems with conversational agents (i.e., talking heads) help students learn by holding conversations in natural language. One version of AutoTutor is sensitive to the emotions of students in addition to their cognitive states. Coh-Matrix analyzes texts on multiple levels of language and discourse, such as text genre, cohesion, syntax, and word characteristics. Coh-Matrix can assist students, teachers, principals, and policymakers when they make decisions on the right text to assign to the right student at the right time. Computers are not perfect conversation partners and comprehenders of text, but the current systems are undeniably useful.

Keywords: discourse processing, learning technologies, emotions, comprehension

Discourse is a general term for written, spoken, and other forms of communication. We are engaged in discourse

when we hold conversations, read texts, write memos, work on the computer, watch television, interpret traffic signs, and use most electronic appliances. Discourse therefore permeates most of our everyday lives in one form or another. Discourse is often captured in language, but there are also many nonlinguistic communication channels, such as facial expressions, gestures, and even body posture. Linguistic messages are sometimes conveyed in print, but oral conversation is the dominant medium most of the time for most people in most cultures.

We have reached a point in history when computers can simulate (or emulate) many aspects of discourse comprehension, generation, and interaction. The vision of humans communicating with computers in natural language has fascinated science fiction writers for decades. This vision has been shifting from science fiction to reality with advances in computational linguistics, corpus linguistics, artificial intelligence, information retrieval, data mining, and discourse processing (Graesser, Gernsbacher, & Goldman, 2003; Jurafsky & Martin, 2008; Landauer, McNamara, Dennis, & Kintsch, 2007; Shermis & Burstein, 2003). There are computer tutors that hold conversations in natural language and help students learn subject matters almost as well as human tutors (Graesser, Jeon, & Dufty, 2008). There are systems that can grade student essays as well as experts in English composition can (Shermis & Burstein, 2003). Computer systems can detect the emotions of learners on the basis of dialogue history, facial expressions, and body posture, with accuracy scores on par with those of humans trained to detect emotions (D'Mello & Graesser, 2010). A system called Linguistic Inquiry Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) can identify the personalities, social status, and other psychological characteristics of writers by classifying the words they use on dozens of psychological categories. Discourse patterns can unveil the characteristics and status of individuals varying from political leaders to terrorists (Hancock et al., 2010). The fusion of psychology with computer science, linguistics, and other fields has not only advanced the science of dis-

Editor's Note

Arthur C. Graesser received the Award for Distinguished Contributions of Applications of Psychology to Education and Training. Award winners are invited to deliver an award address at the APA's annual convention. A version of this award address was delivered at the 119th annual meeting, held August 4–7, 2011, in Washington, DC. Articles based on award addresses are reviewed, but they differ from unsolicited articles in that they are expressions of the winners' reflections on their work and their views of the field.

course but has also provided learning environments and useful text analysis tools.

My career has been consumed with understanding discourse structure and processes. I have always believed that discourse is just as fundamental as perception, attention, memory, learning, decision making, problem solving, personality, social interaction, and other core faculties that have traditionally received top billing in psychology textbooks and curricula. All of these faculties are significantly shaped by discourse. Consider memory, which everyone agrees is fundamental. What we remember about an event is often constrained and explained by the surrounding discourse. For example, memory for an event in a story is influenced by the event's significance in the surrounding plot. A remark at a party is memorable when it is construed as an insult in the broader conversational context. A fact is remembered from an editorial when it supports an argument that persuades the reader. Perhaps introductory psychology textbooks in the future will have a chapter on discourse and communication.

This article describes discourse technologies that automatically analyze discourse, language, and emotions in order to help people learn and think at deeper conceptual levels. The first section covers computer agents (primarily AutoTutor) that help students learn through conversation, whereas the second section focuses on agents that are sensitive to student emotions. The third section describes a system (Coh-Metrix) that analyzes texts on multiple levels of discourse. The technologies were developed by my colleagues and I at the interdisciplinary Institute for Intelligent Systems at the University of Memphis, as well as by colleagues at other institutions who are contributing to an emerging *computational discourse science*. These efforts help us better understand discourse mechanisms in addition to helping students think and learn.

Learning and Thinking With Conversational Agents

The technologies of interest have computer agents that hold conversations with students in natural language. The content of what the agent expresses is designed to emulate aspects of human tutoring that are known to help students learn. Conversation and social interaction are the centerpieces of several educational theories (Resnick, 2010; Vygotsky, 1978), so this approach would presumably be beneficial for computer tutors.

This section focuses on AutoTutor (Graesser, Jeon, & Dufty, 2008; Graesser, Lu, et al., 2004), but other noteworthy systems include ITSPOKE (Litman et al., 2006), Tactical Language and Culture Training System (Johnson & Valente, 2008), Why-Atlas (VanLehn et al., 2007), Operation ARIES! (Millis et al., in press), and iSTART (McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). Collectively, these systems help students learn a variety of subject matters and skills, such as computer literacy, elec-

tronics, physics, circulatory systems, critical thinking about science, foreign languages, cultural practices, and reading strategies. These systems attempt to interpret the student's verbal contributions and generate *discourse moves* (e.g., verbal or nonverbal responses, sentences) that both advance coverage of the material and adjust to what the system believes the student knows. These systems are not rigid scripted lectures.

AutoTutor has significantly helped students learn in dozens of experiments that target the areas of computer literacy and conceptual physics. The system shows learning gains of approximately 0.80σ (standard deviation units) compared with pretests or with a condition that has students read a textbook for an equivalent amount of time (Graesser, Lu, et al., 2004; VanLehn et al., 2007). It is most effective for deeper conceptual levels of comprehension and reasoning (e.g., why, how, what if), as opposed to shallow facts (e.g., who, what, when, where). AutoTutor's impact on learning compares favorably with that of human tutors. For example, learning gains are approximately 0.4σ for typical unskilled tutors in the school systems, when compared to classroom controls and other suitable controls (Cohen, Kulik, & Kulik, 1982), and learning gains vary from 0.2 to 2.0σ for accomplished human tutors (Bloom, 1984; Chi, Roy, & Hausmann, 2008; VanLehn et al., 2007).

One possible reason that AutoTutor helps learning is because its dialogue mechanisms incorporate discourse patterns of human tutors. Prior to the development of AutoTutor, my colleagues and I conducted a series of studies that analyzed the language and discourse of naturalistic human tutoring in fine detail (Graesser, D'Mello, & Person, 2009; Graesser & Person, 1994). Many of these discourse mechanisms were incorporated in AutoTutor. However, AutoTutor also includes more ideal strategies that are rarely exhibited by human tutors, and these also help learning. For example, human tutors typically summarize a learning segment for the student after it is finished, whereas a more ideal strategy would be to request that the student supply the summary.

We have not entirely pinned down what aspects of AutoTutor's discourse are responsible for the learning. However, we do know that some aspects of AutoTutor are *not* good candidates for explaining the learning gains. AutoTutor has an animated conversational agent (talking head) with speech, facial actions, facial expressions, and some hand gestures. Comparisons have been made between AutoTutor versions with pure text and versions that vary the presence/absence of the agent's speech and/or facial expressions. The text versions are nearly as effective as a full-blown animated conversational agent. Analogously, we have compared versions in which the students type in their contributions via keyboard and versions with spoken student input via commercial speech recognition systems. The typed and spoken input versions yield similar learning

gains, but there is a slight advantage for the typed input version because the spoken version has speech recognition errors (D'Mello, Dowell, & Graesser, 2011). Therefore, it is the content of what gets said that most matters, not the face or verbal communication medium. Another feature that is not particularly effective consists of lengthy stretches of didactic information, such as reading pages of text from a textbook or listening to a spoken lecture. Instead, it is the conversational interactivity that matters, with tutor contributions adapting to student contributions.

Given the importance of conversational content, how does AutoTutor manage the discourse? AutoTutor's tutorial dialogues are organized around difficult questions and problems that require reasoning and explanations in the answers. The following are examples of challenging questions on the topics of Newtonian physics and computer literacy.

Physics question: If a lightweight car and a massive truck have a head-on collision, which vehicle undergoes the greater change in its motion, and why?

Computer literacy question: When you turn on the computer, how is the operating system first activated and loaded into RAM?

These why and how questions require from three to seven sentences in an ideal answer. However, students rarely express more than a couple of sentences when initially asked these deep questions. It takes a conversation, typically 20 to 100 turns, to draw out more of what the student knows and to answer the questions collaboratively.

AutoTutor has a number of *dialogue moves* when it constructs a conversational turn and manages the collaborative dialogue in a fashion that encourages more student contributions:

1. *Short feedback* on the quality of the contribution in the student's previous turn, such as positive ("very good") or neutral ("okay") versus negative ("not quite").
2. *Pumps* encourage the student to express more information ("What else?").
3. *Hints* guide the student to express sentence-length ideas that are important answers to the main question or problem. For example, the hint "What about the forces of the vehicles on each other?" attempts to get the student to express, "The forces exerted by each vehicle on each other are equal in magnitude."
4. *Prompts* guide the student to fill in a missing word in an important idea. To get the student to express the word "magnitude," for example, AutoTutor would deliver the prompt "The forces of the two vehicles on each other are equal in *what*?"
5. *Assertions* of AutoTutor articulate important ideas in the answer or problem, for example, "The

forces of the two vehicles on each other are equal in magnitude."

6. *Corrections* correct erroneous ideas and misconceptions. After the student expresses the misconception "The smaller vehicle exerts less force on the larger vehicle," then AutoTutor corrects the student with the assertion in #5 above.
7. *Answers* are provided when the students ask some types of questions, such as definitional questions (e.g., "What does acceleration mean?"). However, students do not frequently ask questions in both human and computer tutoring sessions because the tutor is prone to drive the agenda.
8. *Summaries* provide the complete answer to the main question or problem.

Most of AutoTutor's conversational turns include two or more of these dialogue moves. For example, after a student expresses a misconception, AutoTutor would have a conversational turn that generates short negative feedback, a correction, and then a hint, as illustrated below.

Student: The smaller vehicle exerts less force on the larger vehicle.

Tutor: No, the forces of the two vehicles on each other are equal in magnitude. What about the velocity of the two vehicles?

The three to seven sentences in a full answer to the main question are eventually constructed. These sentences are parts of an explanation that captures important principles of the subject matter.

One important goal is to get the student to express the answer because the active generation of an explanation is better than passive learning. AutoTutor measures and tracks the extent to which the students' verbal contributions match good answers to the question (called *expectations*) versus bad answers (called *misconceptions*). Students receive higher scores to the extent that they express more of the expectations and fewer of the misconceptions in the tutorial dialogue. However, AutoTutor cannot interpret student contributions that have no matches to the anticipated expectations and misconceptions; it can only make comparisons between the student input and anticipated ideas through *pattern matching algorithms*. Interestingly, human tutors tend to be similarly constrained. Human tutors also have trouble understanding and responding to student answers that are substantially outside of the scope of expected content (Graesser, D'Mello, & Person, 2009).

Student contributions rarely match the expectations perfectly because natural language tends to be imprecise, fragmentary, vague, and ungrammatical. AutoTutor implements semantic match algorithms that can accommodate the scruffiness of natural language (Graesser, Penumatsa, Ventura, Cai, & Hu, 2007; Rus, McCarthy, McNamara, & Graesser, 2008). These semantic match algorithms are computed on individual student turns, combinations of

turns, or the cumulative sequence of turns that lead up to a particular point in the dialogue.

AutoTutor generates dialogue moves to fill in missing content and achieve *pattern completion*. The system periodically identifies a missing expectation during the course of the dialogue and posts the goal of covering the expectation. When a particular expectation is posted, AutoTutor tries to get the student to express it by generating hints and prompts that encourage the student to fill in missing ideas and words. The selection of AutoTutor's dialogue moves is sensitive to the cognitive states of the learner. For example, students who have more knowledge and verbal abilities provide most of the information in the answer, so AutoTutor generates primarily pumps and hints. In contrast, students with low knowledge and/or verbal abilities need more prompts and assertions from AutoTutor (Graesser et al., 2007). There is a continuum from the student to the tutor supplying information as the system moves from pumps, to hints, to prompts, to assertions.

The conversations managed by AutoTutor are not always perfectly smooth, but they do help the students learn, and the dialogue is adequate for students to get through the sessions with minimal irregularities. Interestingly, it is difficult for judges to decide whether the content of a particular turn in the dialogue was generated by AutoTutor or by an expert human tutor of computer literacy (Person, Graesser, & the Tutoring Research Group, 2002). Person, Graesser, and the Tutoring Research Group randomly sampled AutoTutor turns and half of the time substituted content generated by human tutors at the sample points in the dialogue. Judges were presented these tutoring turns in a written transcript and were asked to decide whether each was generated by a computer or a human. The judges were not able to significantly discriminate whether the particular turns were generated by humans or AutoTutor. Nevertheless, judges and learners themselves are presumably able to decide whether a *sequence* of turns is part of a dialogue with AutoTutor or with a human tutor.

The errors and limitations of AutoTutor are just as illuminating about discourse mechanisms as are its successes. The limitations below are currently being investigated in our current discourse technologies:

1. AutoTutor sometimes makes errors in evaluating the quality of student contributions, so AutoTutor's short feedback is incorrect and the student gets confused or frustrated.
2. AutoTutor makes errors in assigning student contributions to the correct speech act category—for example, question, assertion, meta-comment (“I’m lost”), so AutoTutor’s response is not relevant and coherent.
3. AutoTutor cannot answer many of the student questions, so the answers do not seem relevant and

students are prone to stop asking questions. However, transcripts of human tutoring show a low frequency of student questions (Graesser & Person, 1994).

4. AutoTutor is limited in its mixed-initiative dialogue because it cannot handle changes in topics, tangents, and off-the-record contributions of students.

Many versions of AutoTutor and derivatives of AutoTutor have been developed since its inception in 1997. Table 1 presents a list of the systems with conversational agents that have been developed in my collaborations with colleagues at the University of Memphis or at other universities; often, these collaborators have taken the lead on these funded projects. The design of all of these systems is grounded in principles of learning that are endorsed by the cognitive and learning sciences (Graesser, Halpern, & Hakes, 2008). Most of these systems have facilitated learning when tested on students in middle school, high school, or college.

Emotions With Conversational Agents

Conversational agents have recently been designed to respond to students' emotions in addition to their cognitive states. An adequate understanding of affect–learning connections is essential to the design of engaging educational artifacts that range from responsive intelligent tutoring systems on technical material to entertaining media and games. Therefore, our designs of AutoTutor and other systems with conversational agents have documented the emotions that learners experience while using these advanced learning environments (Baker, D’Mello, Rodrigo, & Graesser, 2010; D’Mello, Craig, & Graesser, 2009). Our recent emotion-sensitive AutoTutor (AutoTutor-ES) automatically detects learner emotions based on multiple channels of discourse (D’Mello & Graesser, 2010) and responds appropriately to the students' affect states by selecting appropriate discourse moves and displaying emotions in facial expressions and speech (D’Mello & Graesser, in press).

The role of emotions in complex learning has been explored in the contexts of human tutoring, classrooms, and other educational contexts (Lepper & Woolverton, 2002; Meyer & Turner, 2006; Pekrun, 2006) as well as in more general cognition activities (Bower, 1992; Mandler, 1984; Ortony, Clore, & Collins, 1988; Russell, 2003). Interestingly, the “universal” emotions that Ekman (1992) investigated (e.g., sadness, happiness, anger, fear, disgust, surprise) have minimal relevance to learning-centered contexts, where the dominant affective states include confusion, frustration, boredom, flow/engagement, delight, and surprise (Baker et al., 2010; D’Mello et al., 2009). The affect state of anxiety also occurs when students are being evaluated.

Table 1
Learning Environments With Conversational Agents Developed by Graesser and Collaborators

System	Brief description	Leader/collaborator
AutoTutor	Conversational tutor on computer literacy and physics	
AutoTutor-3D	Physics with embedded interactive simulation in 3D world	Tanner Jackson
AutoTutor-Lite	Simplified discourse applied to Powerpoint on any topic	Xiangen Hu
AutoTutor-ES	AutoTutor being sensitive to learners' emotions	Sidney D'Mello
AutoMentor	Multiparty serious game with mentor on urban planning	David Shaffer
DeepTutor	Physics tutor with deep natural language processing	Vasile Rus
GnuTutor	Open source version of AutoTutor on any topic	Andrew Olney
GuruTutor	Biology tutor with deep natural language and gesture	Andrew Olney
HURAA Advisor	Web tutor on ethical treatment of subjects in experiments	Xiangen Hu
iDRIVE	Learning to ask deep questions on science topics	Barry Gholson and Scotty Craig
iSTART, iSTART-ME	Learning to generate self-explanations while reading text	Danielle McNamara
MetaTutor	Learning skills of self-regulated learning and metacognition	Roger Azevedo
Operation ARIES!	Critical reasoning about scientific methods	Keith Millis, Diane Halpern, and Zhiqiang Cai
Writing-Pal	Learning to write argumentative essays	Danielle McNamara

The cognitive-affective state of confusion is particularly interesting because it theoretically is expected to play an important role in learning and empirically has a positive correlation with learning gains (D'Mello et al., 2009; Graesser et al., 2009). Confusion is diagnostic of *cognitive disequilibrium*, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Festinger, 1957; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Piaget, 1952). Cognitive equilibrium is restored after thought, reflection, problem solving, and other effortful cognitive activities. Cognitive disequilibrium is a critical juncture in the learning process that is sensitive to individual differences. Some students give up when experiencing confusion because they have a self-concept that they are not good at the subject matter or because they prefer not to receive negative feedback (Dweck, 1999; Meyer & Turner, 2006). Other students treat confusion as a challenge to conquer and expend cognitive effort to restore equilibrium. The first type of student needs encouragement, hints, and prompts to get him or her over the hurdle, whereas the second type would best be left to his or her own devices. An adaptive tutor would treat these students differently.

AutoTutor-ES responds to different profiles of the students' emotional and cognitive states (D'Mello & Graesser,

in press). If the learner is frustrated, for example, the tutor gives hints or prompts to advance the learner in constructing knowledge and makes supportive empathetic comments to enhance motivation. If the learner is bored, the tutor presents more engaging material or challenging problems for the more knowledgeable learner. The tutor continues business as usual when the learner is in a state of flow (Csikszentmihalyi, 1990), that is, when the learner is so deeply engaged in learning the material that time and fatigue disappear. The emotions of delight and surprise are fleeting, so there is no need to respond to these states in any special way. AutoTutor's intervention when the student is confused is both critical and complex, as previously discussed. One speculation is that each student has a *zone of optimal confusion* that varies with his or her background knowledge and interest in the subject matter.

An automated emotion classifier is necessary for AutoTutor-ES to be responsive to learner emotions. We have developed and tested an automated emotion classifier for AutoTutor based on the dialogue history, facial action units, and position of the student's body during tutoring (D'Mello, Dale, & Graesser, 2011; D'Mello & Graesser, 2010). There are systematic relations between these sensing channels and particular emotions. With respect to dialogue history, emotions are predicted by (a) the occurrence of

AutoTutor's feedback, (b) the type of feedback (positive, neutral, negative), (c) the directness of AutoTutor's dialogue moves (e.g., hints are less direct than assertions), (d) the quality of the learner's contributions, and (e) the phase of the tutoring session (early versus late). Regarding the nonverbal channels, emotions are correlated with particular facial expressions, posture, and face-posture-dialogue combinations. Confusion, surprise, and delight are most directly manifested by facial expressions, whereas frustration is best predicted by dialogue history, and posture dynamics are needed to discriminate boredom, engagement/flow, and neutral states. AutoTutor's body pressure measurement system has revealed that bored students either fidget or have a large distance between their face and the screen. The features from the various modalities can be detected in real time automatically on computers, so we have integrated these sensing technologies with AutoTutor-ES.

It is too early to make any firm conclusions about the impact of AutoTutor-ES on learning, but we have conducted some studies. We have compared the original AutoTutor without emotion tracking to an AutoTutor version that is emotionally supportive. The supportive AutoTutor would have polite and encouraging positive feedback ("You're doing extremely well.") or negative feedback ("This is difficult for most students."). There is another version that tries to shake up the emotions of the student by being playfully rude and telling the student what emotion the student is having ("I see that you are frustrated."). Instead of giving earnest feedback, the rude AutoTutor gives positive feedback that is sarcastic (e.g., "Aren't you the little genius?") and negative feedback that is derogatory (e.g., "I thought you were bright, but I sure pegged you wrong."). The simple substitution of this feedback dramatically changes AutoTutor's personality. The rude tutor is very engaging for some students, whereas other students would prefer to interact with the polite, supportive tutor. The data we have collected reveal that the impact on learning appears to depend on the phase of tutoring and the student's level of mastery. An emotion-sensitive AutoTutor had either no impact or a negative impact on learning during early phases of the tutoring session. During the later stages, the polite, supportive AutoTutor improved learning, but only for the low-knowledge students. Although more studies need to be conducted, it is tempting to speculate (a) that emotional displays by AutoTutor may not be beneficial during the early phases of an interaction when the student and the agent are "bonding," (b) that a supportive, polite tutor is appropriate at later phases for students who have low knowledge and abilities, and (c) that the playful, rude tutor is motivating when boredom starts emerging for the more confident, high-knowledge learners.

Emotions are of course central to the design of educational games (Conati, 2002; McNamara, Jackson, & Graesser, in press; Millis et al., in press; Moreno & Mayer,

2005; Shaffer, 2006). Educational games ideally are capable of turning work into play by minimizing boredom, optimizing engagement/flow, presenting challenges that reside within the optimal zone of confusion, preventing persistent frustration, and engineering delight and pleasant surprises.

Automated Analysis of Discourse at Multiple Levels

Another approach to helping students learn is to assign texts that fit the student's profile of reading proficiency. The computer could be used to assign the right text to the right student at the right time. For example, the selected text might push the envelope at the reader's challenge zone, based on what the computer tracks about the reader and about the text. The computer algorithm would consider cognitive, motivational, emotional, and social characteristics of the reader. Readers with high self-efficacy (i.e., they are convinced they can perform well) and a preference for 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 avoidance of academic risk taking would receive texts within their challenge zone.

In this section I shift from tutorial dialogue to automated analyses of text and other types of discourse (such as speeches and plays). My focus is on an automated text analysis system called Coh-Metrix (Graesser & McNamara, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, McCarthy, & Graesser, 2010; see <http://cohmetrix.memphis.edu>). Coh-Metrix is a computer facility that analyzes text on multiple levels of discourse and language. When the user enters text into a web facility, Coh-Metrix produces dozens of measures of cohesion, syntax, and words. This permits researchers and practitioners to examine characteristics of text that go beyond the traditional metrics of readability and text difficulty, such as Flesch-Kincaid age levels (Klare, 1974–1975), Lexile scores (Stenner, 2006), and Degrees of Reading Power scores (DRP; Koslin, Zeno, & Koslin, 1987). Word frequency, word length, and sentence length are strong predictors of these simple, single-dimension metrics of text difficulty, whereas Coh-Metrix captures many more levels and characteristics of text.

Psychological theories of comprehension have identified the representations, structures, strategies, and processes at multiple levels of language and discourse (Clark, 1996; Graesser & McNamara, 2011; Kintsch, 1998; Snow, 2002). These multilevel frameworks have typically proposed the following levels: *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 organization), and the *pragmatic communication* level (between speaker and listener or be-

tween writer and reader). These six levels are briefly elaborated in Table 2.

The meaning of most of the levels can be readily reconstructed from Table 2. However, a bit more should be conveyed about the textbase, situation model, genre, and rhetorical structure because they are particularly relevant to cohesion—the hallmark of Coh-Metrix. The textbase contains explicit ideas in the text in a form that preserves the meaning but not the precise wording and syntax (Kintsch, 1998). Referential cohesion is achieved in the textbase to the extent that explicit noun phrases in sentences refer to explicit words and ideas in other sentences. The situation model is the subject matter that is being described in an informational text or the world that evolves in a story (Kintsch, 1998; O'Brien, Rizzella, Albrecht, & Halleran, 1998; Van den Broek, Rapp, & Kendeou, 2005; Zwaan & Singer, 2003). For example, the situation model of a story includes the people, objects, spatial setting, actions, events, processes, plans, thoughts, and emotions of people, and other content at deeper levels of meaning. Situation model cohesion is achieved to the extent that the content can be conceptually related at these deeper levels or that there are connectives (e.g., *because, in order to, therefore, later on*) that stitch together potential cohesion gaps. The text genre is the type of discourse, such as a news story, a novel, a persuasive sermon, or a science text that explains a causal mechanism. The rhetorical structure is the global organiza-

tion of the text and the discourse function of particular excerpts. Coh-Metrix measures texts on these first five levels but not on the pragmatic communication level that is constrained primarily by the context of use rather than the text per se.

It is beyond the scope of this article to describe the computational mechanisms that analyze texts at these different levels. A few highlights should offer a glimpse of what is involved when generating measures at the different levels. The measures are grouped into those related most to words, sentence structure, and connections between sentences.

Words

The quality of a person's lexicon is critical for understanding text at the various levels of discourse (Perfetti, 2007). The word characteristics include many categories of parts of speech (e.g., nouns, verbs, adjectives, adverbs, prepositions, pronouns). Connectives are an important category because they are important for establishing situation model cohesion, as in the case of causal (*because, so*), temporal (*and then, after, during*), additive (*also, moreover*), and adversative (*on the other hand, however*) connectives (Louwerse, 2001). Word frequency in Coh-Metrix is computed as the occurrence frequency (per million words) of a word appearing in a representative set of published documents. Word length has a very strong negative correlation with word frequency and serves as an excellent proxy for world

Table 2
Levels of Discourse and Language

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	Grammatical structure (noun phrase, verb phrase, clause) Linguistic style
Textbase	Semantic meaning of explicit propositions or clauses Words linked to other explicit text constituents
Situation model	Situation conveyed in the text (people, objects, spatial layout, events) 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 (old) versus new information Images and mental simulations of events
Genre and 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)

knowledge; syllables per word is indeed one of the two major components of the Flesch-Kincaid readability formula. Coh-Metrix measures words on characteristics in an established psycholinguistic database (the MRC Psycholinguistic Database; Coltheart, 1981), a collection of human ratings of several thousands of words along several psychological dimensions: age of acquisition, meaningfulness, concreteness, imagability, and familiarity. The semantic content of words is analyzed by WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), a lexicon with semantic dimensions, change-of-state verbs, animate nouns, and polysemy (multiple word senses).

Sentences

The sentence measures compute information load (words per sentence) and syntactic composition. The simplest measure is the number of words per sentence, one of the two major parameters of the Flesch-Kincaid grade level and an obvious measure of working memory load. More subtle measures of syntactic complexity assign syntactic structures to sentences. Syntactic complexity increases with more modifiers per noun phrase, more words before the main verb of the main clause, and the occurrence of sentences in the passive voice rather than the active voice.

Connections Between Sentences

These measures are particularly relevant to the cohesion of the textbase and situation model. Some metrics involve pairs of adjacent sentences in the text, whereas others span all pairs of sentences within each paragraph. Coh-Metrix tracks different types of word co-reference, such as content word overlap, noun overlap, and stem overlap. *Content word overlap* is the proportion of content words that are the same between pairs of sentences. *Noun overlap* is the proportion of all sentence pairs that share one or more common nouns, whereas *stem overlap* is the proportion of sentence pairs in which a word in one sentence has the same semantic morpheme as a word in the other sentence (e.g., the noun “invasion” and the verb “invaded”). Connectives are important words for relating the events, actions, and states expressed in the text with respect to cohesion of the situation model.

Coh-Metrix measures conceptual overlap between sentences with a statistical model of word meaning called *latent semantic analysis* (LSA; Landauer et al., 2007). LSA is an important method of computing similarity because it considers implicit knowledge in addition to the explicit words. LSA is a mathematical, statistical technique for representing word and world knowledge and is based on a large corpus of texts. The conceptual similarity between any two text excerpts (e.g., word, clause, sentence, text) is computed as a statistical overlap in the values and weighted dimensions of the words in the two text excerpts. LSA-based cohesion is measured between adjacent sen-

tences, between all pairs of sentences in a paragraph, and also by how much old versus new information occurs in successive sentences in the text.

Nearly 1,000 measures of language and discourse have been computed with Coh-Metrix over the last decade. Nevertheless, we have discovered that a small number of factors (principal components) robustly account for text variations in analyses of 37,520 texts (Graesser, McNamara, & Kulikowich, 2011). Principal-components analyses have shown that the following factors account for approximately 67% of the variability among texts:

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 explicit words and ideas that overlap across sentences and the text.
3. *Situation model cohesion*: Causal, intentional, and temporal connections help the reader 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 understand. Complex sentences have structurally embedded syntax.
5. *Word concreteness*: Concrete words evoke mental images and are more meaningful to the reader than are abstract words.

The above dimensions are aligned with the theoretical components in Table 2. This is a remarkable convergence of empirical data with the multilevel theoretical framework.

Coh-Metrix can lend a hand in improving literacy in a number of ways. The text characteristic profiles (i.e., scores on the five factors or the more specific measures) can be generated automatically by the computer for any text. This opens the door for an automated selection of texts that assigns the right text to the right student at the right time. Texts can be selected to help remediate deficits at particular theoretical levels. Another approach to improving literacy is to collect reading time or comprehension data from particular readers on texts that have different discourse characteristics. Such data can help diagnose reading problems at particular discourse levels and remediate the problems with appropriate interventions. If a reader has trouble with texts that have complex syntax, for example, then the reader might benefit from training that is targeted for syntax. If a reader has trouble comprehending narrative texts with temporal discontinuities (e.g., flashbacks and flash-forwards), then the reader would presumably benefit from training on stories with time transformations. Readers who have trouble comprehending science texts would benefit from a combination of remedial techniques: instruction on relevant vocabulary and core mental models as well as the

self-explanation training provided by systems such as iSTART (McNamara et al., 2007). Individualized student instruction is more effective than having all students in the curriculum move at the same pace on the same materials (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Lesgold & Welch-Ross, 2011).

Comprehension performance depends on the characteristics of the test questions in addition to the texts. Automated tools are needed to analyze the discourse characteristics of the test questions and to relate them to the texts that the readers read. One way to do this is to enter the test questions into Coh-Metrix and identify possible difficulties with the questions. However, test questions are normally short and have unusual formatting constraints that do not mesh well with Coh-Metrix. A second approach is to submit the test questions to QAID (Question Understanding Aid), a web tool that automatically critiques questions on the difficulty of the words, syntactic complexity, and working memory load (Graesser, Cai, Louwerse, & Daniel, 2006). One useful tool to be developed in the future is a question analysis *workbench* that systematically analyzes questions that are associated with (a) texts at different discourse levels and (b) subject matter content at different levels of complexity (Graesser, Ozuru, & Sullins, 2009).

In closing, in this article I have made the case for a computational discourse science that integrates contributions in the fields of psychology, computer science, linguistics, and discourse processing. The tools developed by this interdisciplinary fusion can advance the educational enterprise by helping students learn and think in ways that are sensitive to their cognitive and emotional states. The tools that my colleagues and I have developed, such as AutoTutor and Coh-Metrix, play a role in advancing science in addition to helping students.

It is important to reiterate that computers are not perfect conversation partners and comprehenders of text. No one seriously believes that a computer could replace a spouse or understand Shakespeare. Current computer systems cannot comprehend humor, sarcasm, tragedy, and sexuality with any modicum of sophistication. Not yet and likely never. However, an imperfect computer system, just like an imperfect human, can sometimes be useful and help students learn.

Author's Note

The research described here was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918, BCS 0904909, ALT-0834847, DRK12-0918409), the Institute of Education Sciences (R305H050169, R305B070349, R305A080589, R305A080594, R305G020018), the Office of Naval Research (N00014-00-1-0600), the Gates Foundation, the U.S. Department of Homeland Security (Z934002/UTAA08-063), 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 author and do not necessarily reflect the views of these funding sources.

Correspondence concerning this article should be addressed to Arthur C. Graesser, Department of Psychology, University of Memphis, 202 Psychology Building, 400 Innovation Drive, Memphis, TN, 38152-3230. E-mail: a-graesser@memphis.edu

References

- Baker, R. S., D'Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68, 223–241. doi:10.1016/j.ijhcs.2009.12.003
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4–16.
- Bower, G. (1992). How might emotions affect learning? In S. A. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 3–31). Hillsdale, NJ: Erlbaum.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32, 301–341. doi:10.1080/03640210701863396
- Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237–248.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 33A, 497–505. doi:10.1080/14640748108400805
- Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence*, 16, 555–575. doi:10.1080/08839510290030390
- Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: Al-

- gorithm-guided individualized reading instruction. *Science*, 315, 464–465. doi:10.1126/science.1134513
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.
- D'Mello, S. K., Craig, S. D., & Graesser, A. C. (2009). Multimethod assessment of affective experience and expression during deep learning. *International Journal of Learning Technology*, 4, 165–187. doi:10.1504/IJLT.2009.028805
- D'Mello, S., Dale, R., & Graesser, A. (2011). Disequilibrium in the mind, disharmony in the body. *Cognition & Emotion*. Advance online publication. doi:10.1080/02699931.2011.575767
- D'Mello, S. K., Dowell, N., & Graesser, A. (2011). Does it really matter whether students' contributions are spoken versus typed in an intelligent tutoring system with natural language? *Journal of Experimental Psychology: Applied*, 17, 1–17. doi:10.1037/a0022674
- D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20, 147–187. doi:10.1007/s11257-010-9074-4
- D'Mello, S., & Graesser, A. (in press). Emotions during learning with AutoTutor. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive technologies for training and education*. Cambridge, England: Cambridge University Press.
- Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169–200. doi:10.1080/02699939208411068
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Graesser, A. C., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, 70, 3–22. doi:10.1093/poq/nfj012
- Graesser, A. C., D'Mello, S. K., & Person, N., (2009). Meta-knowledge in tutoring. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 361–382). New York, NY: Routledge.
- Graesser, A. C., Gernsbacher, M. A., & Goldman, S. (Eds.). (2003). *Handbook of discourse processes*. Mahwah, NJ: Erlbaum.
- Graesser, A. C., Halpern, D. F., & Hakel, M. (2008). 25 principles of learning. Washington, DC: Task Force on Lifelong Learning at Work and at Home, Association for Psychological Science. Retrieved from <http://psyc.memphis.edu/learning/whatwewknow/index.shtml>
- Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, 45, 298–322. doi:10.1080/01638530802145395
- 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. doi:10.3758/BF03195563
- 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. doi:10.3758/BF03193225
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371–398. doi:10.1111/j.1756-8765.2010.01081.x
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223–234. doi:10.3102/0013189X11413260
- 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. doi:10.3758/BF03195564
- Graesser, A., Ozuru, Y., & Sullins, J. (2009). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (pp. 112–141). New York, NY: Guilford Press.
- Graesser, A. C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed initiative dialogue in natural language. In T. Lan-

- dauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 243–262). Mahwah, NJ: Erlbaum.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137. doi:10.3102/00028312031001104
- Hancock, J. T., Beaver, D. I., Chung, C. K., Frazee, J., Pennebaker, J. W., Graesser, A., & Cai, Z. (2010). Social language processing: A framework for analyzing the communication of terrorists and authoritarian regimes. *Behavioral Sciences of Terrorism and Political Aggression, 2*, 108–132. doi:10.1080/19434471003597415
- Johnson, L. W., & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In M. Goker & K. Haigh (Eds.), *Proceedings of the Twentieth Conference on Innovative Applications of Artificial Intelligence* (pp. 1632–1639). Menlo Park, CA: AAAI Press.
- Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Englewood Cliffs, NJ: Prentice Hall.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Klare, G. R. (1974–1975). Assessing readability. *Reading Research Quarterly, 10*, 62–102. doi:10.2307/747086
- Koslin, B. I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York, NY: College Entrance Examination Board.
- Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). (Eds.). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- Lepper, M., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135–158). Orlando, FL: Academic Press doi:10.1016/B978-012064455-1/50010-5
- Lesgold, A., & Welch-Ross, M. (2011). Learning sciences: Foundations and applications to adolescent and adult literacy [Manuscript under revision]. Washington, DC: National Academy of Sciences.
- Litman, D. J., Rose, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education, 16*, 145–170.
- Louwerse, M. M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics, 12*, 291–315. doi:10.1515/cogl.2002.005
- Mandler, G. (1984). *Mind and body: The psychology of emotion and stress*. New York, NY: W. W. Norton.
- McNamara, D. S., Jackson, G. T., & Graesser, A. (2010). Intelligent tutoring and games (ITaG). In Y. K. Baek (Ed.), *Gaming for classroom-based learning: Digital role-playing as a motivator of study* (pp. 44–65). New York, NY: IGI Global.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292–330. doi:10.1080/01638530902959943
- 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.
- Meyer, D. K., & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377–390. doi:10.1007/s10648-006-9032-1
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*, 235–244. doi:10.1093/ijl/3.4.235
- Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (in press). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.), *Serious games and edutainment applications*. London, England: Springer-Verlag.
- Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*, 117–128. doi:10.1037/0022-0663-97.1.117
- O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1200–1210. doi:10.1037/0278-7393.24.5.1200

- Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. New York, NY: Cambridge University Press.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. doi:10.1007/s10648-006-9029-9
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC 2007 [Computer software]. Austin, TX: LIWC Inc.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383. doi:10.1080/10888430701530730
- Person, N. K., Graesser, A. C., & the Tutoring Research Group. (2002). Human or computer? AutoTutor in a bystander Turing test. In S. A. Cerri, G. Gouarderes, & F. Paraguaçu (Eds.), *Intelligent tutoring systems 2002* (pp. 821–830). Berlin, Germany: Springer.
- Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press. doi:10.1037/11494-000
- Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher, 39*, 183–197. doi:10.3102/0013189X10364671
- Rus, V., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2008). A study of textual entailment. *International Journal on Artificial Intelligence Tools, 17*, 659–685. doi:10.1142/S0218213008004096
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172. doi:10.1037/0033-295X.110.1.145
- Shaffer, D. W. (2006). *How computer games help children learn*. New York, NY: Palgrave Macmillan. doi:10.1057/9780230601994
- Shermis, M. D., & Burstein, J. (2003). (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. Hillsdale, NJ: Erlbaum.
- Snow, C. B. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND.
- Stenner, A. J. (1996, October 31). *Measuring reading comprehension with the Lexile framework*. Paper presented at the California Comparability Symposium, Burlingame, CA Retrieved from http://www.lexile.com/m/resources/materials/Stenner_Measuring_Reading_Comprehension.pdf
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*, 299–316. doi:10.1207/s15326950dp3902&3_11
- 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. doi:10.1080/03640210709336984
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Zwaan, R. A., & Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83–121). Mahwah, NJ: Erlbaum.

