Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Chasing Theory with Technology: A Quest to Understand Understanding

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
McNamara, Danielle S.		

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 02 Aug 2021

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Discourse Processes, Volume 58, Issue 5-6

DOI or URL to published work (if available) https://doi.org/10.1080/0163853X.2021.1917914

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** National Center for Educat through **[Grant number]** R305A1900 to **Institution]** Arizona State U. The opinions expressed are those of the authors and do not represent views of the **[Office name]** National Center for Educa or the U.S. Department of Education.

Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Chasing Theory with Technology: A Quest to Understand Understanding

Danielle S. McNamara

To cite this article: Danielle S. McNamara (2021) Chasing Theory with Technology: A Quest to Understand Understanding, Discourse Processes, 58:5-6, 422-448, DOI: 10.1080/0163853X.2021.1917914

To link to this article: https://doi.org/10.1080/0163853X.2021.1917914

Published online: 02 Aug 2021.

DISCOURSE PROCESSES
2021, VOL. 58, NOS. 5–6, 422–448
https://doi.org/10.1080/0163853X.2021.1917914

Routledge
Taylor & Francis Group

Chasing Theory with Technology: A Quest to Understand Understanding

Danielle S. McNamara

Department of Psychology Arizona State University

ABSTRACT
An overarching motivation driving my research has been to further our theoretical understanding of how readers successfully comprehend challenging text. This article describes the theoretical origins of this research program and my quest to understand comprehension processes through the use of technology. Coh-Metrix was developed to measure, and in turn facilitate, manipulations of text cohesion and text ease. iSTART was developed to provide students with instruction and practice on how to explain text and more effectively make use of limited prior knowledge. In addition, we have developed technologies to measure and change writing quality, relations between ideas, and emerging text comprehension. More recently, our attention has turned to comprehension of multiple documents. Understanding relations between documents and how comprehension emerges when reading multiple sources is important educationally and socially, where the internet provides a continuous stream of reliable and unreliable sources. Across these topics, my collaborators and I have conducted numerous experimental studies, but a central theme to my work has been the use of technology. This article describes these technologies, including natural language processing, game-based tutoring systems, and computational simulations; how they were informed by theory; and how they have informed my theoretical and practical understandings of language, comprehension, social interactions, and cognition as multilayered and multidimensional within what I refer to as the M&M Framework.

Introduction

I remember distinctly sitting in my office at Old Dominion University. It was around 1996. I was outlining some theoretical predictions, hovered over my side desk with stacks of papers and books (the only one without computers on it). My earliest training in Cognitive Psychology had been in 1988 to 1989 with Marilyn Turner, a student of Randy Engle. With her, I conducted experiments and eventually my thesis on working memory. I admit that I was convinced at the start of this endeavor that performance was a function of the size of a box in your brain, but I at least believed there might be *multiple boxes*. The title of my thesis was *The theory of a unitary working memory reexamined: Are there verbal and spatial working memory systems*? I can assure you that I didn't come close to answering that question, but I did learn quite a bit about working memory, mainly because I talked to the participants afterward, and I became convinced that it was not the size of the box(es) but rather strategies and skills that were key to their performance. Seven years later, after training with Alice Healy on cognitive skill training and Walter Kintsch on comprehension, I was even more convinced. I was convinced that working memory was *not* the key to understanding reading, comprehension, problem solving, and other skills. By 1996, I was convinced that the use of strategies was key to skilled

CONTACT Danielle S. McNamara Dsmcnamara1@gmail.com Department of Psychology, Arizona State University, 950 S. McAllister, Tempe, AZ 85287-1104

performance. But, testing these assumptions was challenged by their quasi-experimental nature: Most studies examining the relative contributions of working memory, skills, and strategies are correlational studies. This methodological weakness led me toward using strategy interventions as experimental manipulations rather than solely conducting correlational studies to address these questions.

My initial studies providing strategy instruction were remarkably successful. However, I was also out of my comfort zone. I quickly learned that intervention studies were much more challenging than I had assumed (*seems easy, right, just give them instruction*). It was around this time that I realized I needed to turn back to my own comfort zone, *technology*. This article describes how the use of technology has helped inform my understanding of text and discourse: comprehension, writing, and communication. The overarching purpose of this article is to tell my story in a document of gratitude to the Society for Text and Discourse for awarding me the Distinguished Scientific Contributions Award, to describe how theories of text and discourse inform the development of applied discourse technologies, and in turn, and symbiotically, how the use of technology informs understanding of text and discourse.

The beginning

My studies on text comprehension began when Walter and Eileen Kintsch adopted me into their lab in 1991. This work was prefaced by a year of working with Stephanie Doane, Walter Kintsch, and Peter Polson (e.g., Doane et al., 1992) examining the effects of knowledge and experience on Unix programming, wherein I was thrown into the depths of cognitive simulations using the Construction-Integration (CI) model of comprehension (Kintsch, 1988).

From this point forward, the use of *cognitive simulation* undergirded a large part of my work, or at least how I think. The CI model is basically a simple, constraint-based connectionist model that cheats by starting at the level of the words and is solely feed-forward (more or less). The surprising outcome is that it works remarkably well in accounting for a large number of findings on comprehension and learning from text (Kintsch, 1998). This success can be contrasted with the lack of success for cognitive models such as the 4CAPS (Just & Varma, 2007), SAM (Raaijmakers & Shiffrin, 1980), or ACT-R (Anderson, 1990) in accounting for comprehension processes, whereas they have been remarkably successful in accounting for cognitive tasks that involve processes such as attention, memory, or decision-making. In essence, cognitive simulations are conducted to implement and test theories, and thus it is not surprising that simulations driven by theories of comprehension and built to emulate comprehension processes are more appropriate than those constructed to simulate cognitive processes underpinning attention and memory.

Understanding text is complex; it is an interaction between the reader, the text, and the task. The CI model provided a basic architecture designed to simulate and in turn test those complex interactions. As such, it has guided many of my assumptions when building technologies for text and discourse.

In that light, let's now consider these two sides of comprehension: the text and the reader. Of course, there are multiple facets to each of these—imagine a diamond, with multiple facets, and two sides (depending on the cut of the diamond of course). We begin with the bottom side of the diamond, ***the text.*** We then consider the top side, *the reader*. Finally, I turn to the diamond cutter, ***the writer***. From there, we turn to the diamond mine wherein I describe my more recent work on situations wherein there are multiple speakers or writers and multiple texts. Finally, I end with a brief description of a Multilayered, Multidimensional (M&M) Framework, which postulates that language production and comprehension is both multilayered and multidimensional. Within each of the following sections, I describe research that I and my colleagues have conducted, the technologies, and in turn, the theories that undergird these emergent technologies.

The text

How can we measure the multiple facets and dimensions of text and discourse? That is a question that has occupied a large portion of my career. This quest was largely inspired by two factors. First, in the 1990s, we lacked (easy) access to automated tools to measure basic aspects of text such as word frequency, word concreteness, syntactic complexity, and so on. Access to these multiple features was essential to understanding text and fundamental to conducting well controlled experiments in the field of text and discourse. Second, there existed no automated measures of text cohesion, a facet of text and discourse that had emerged as fundamental to comprehension.

Text cohesion refers to the amount and quality of overlap between ideas in text and discourse. Overlap can occur in terms of explicit words (e.g., nouns, verbs), implied words (anaphor), semantically related words, semantically related ideas, and the underlying parts of speech (i.e., parts of speech, syntactic overlap). When there is greater overlap, text is easier to understand. This is because cohesion gaps require the reader to make inferences to connect the words and ideas in the text. The prior knowledge of the reader plays a large role in the degree to which cohesion affects comprehension. If the reader has little knowledge of the world or the domain, then the negative effects of low cohesion text can be profound. Younger readers and low knowledge readers struggle to comprehend, let alone learn, from text with cohesion gaps (Best et al., 2005; McNamara et al., 1996; McNamara, 2001, 2011; McNamara & Kendeou, 2011).

However, we also know that making learning easy is not effective in the long run. I knew from my graduate studies with Alice Healy about the benefits of generating responses and multiple tests on skill acquisition and learning (McNamara & Healy, 1995a, 1995b). We extended research on the generation effect (which at the time was limited to episodic memory) to the benefits of generating to learn vocabulary and improve math skills (McNamara, 1995). In essence, generating a response (similar to repeated testing) is crucial to learning. In the same way, generating inferences while reading is crucial to learning from text. When a reader has more knowledge about a topic, high cohesion text can stifle readers' need to generate inferences and reduce deep comprehension (McNamara, 2001; McNamara & Kintsch, 1996; O'Reilly & McNamara, 2007; Ozuru et al., 2009).

Tracking and manipulating these complex effects, from multiple types of text cohesion, in the 1990s required laborious discourse analysis, which rendered large-scale analyses virtually impossible. Thus, one of our first undertakings was to develop the tools necessary to conduct large scale text analysis inspired by theories of text and discourse.

Technology: Coh-Metrix

Coh-Metrix is an automated workbench for the analysis of textual features related to text difficulty. I had imagined parts of Coh-Metrix, particular cohesion indices, for many years before the good fortune of my job interview at the University of Memphis in 2002. There I met Max Louwerse and Art Graesser who had the necessary facilities and programmatic foundations, combined with complementary visions of what we needed to build. The Coh-Metrix grant proposal was one of those that we say "wrote itself." Within months of my job interview, we completed the proposal and submitted it, and it was funded that year. Of course, I accepted their offer to join the Cognitive Science program at the University of Memphis, and we began our decade-long collaboration in building Coh-Metrix.

Theory. How did theory drive Coh-Metrix? It might seem to some that Coh-Metrix is a kitchen sink of linguistic indices—it is not. Coh-Metrix was inspired, guided, and carefully aligned with theories of text and discourse, in particular two theories: the CI model (Kintsch, 1998) and the Constructionist theory (Graesser et al., 1994). The former emphasized levels of comprehension and the role of prior knowledge in constructing understandings. The latter emphasized the importance of readers' goals, search for meaning, and rhetorical functions of text (McNamara & Magliano, 2009). As such, while we included basic indices such as number of words, parts of speech, and so on, our ultimate goal was to explain and predict text difficulty guided by theories of comprehension (Graesser & McNamara, 2011). This involved multiple steps of adding indices that were relevant to predicting parts of the theories or simply exploring various aspects of text (McNamara et al., 2014) (Table 1). Each index and set of indices required weeks of debates in our lab meetings on whether to include it, how to implement it, and potential parameters or variations. This was followed by programming the indices, creating

Table 1. Examples of Coh-Metrix measures: Coh-Metrix includes multiple types of indices that describe language generally or are theoretically related text difficulty

Word measures	Referential cohesion
• Number of syllables	• Noun and argument overlap
• Part of speech (noun, verb, pronoun)	• Stem overlap (e.g., *run, runner*)
• Word frequency	• Lexical diversity (e.g., type-token ratio)
• Concreteness, imagery	• Pronominal overlap
• Multiple word meanings	
Syntax	Situation model cohesion
• Structural complexity	• Connectives and discourse markers
• Modifiers per noun phrase	• Causal and intentional verbs
• Words before main verb of main clause	• Causal and intentional cohesion
• Syntactic similarity between sentences	• Repetition in tense and aspect
	• Logical operators (*and, or, therefore, if, then, not*)

corpora, and testing and validating the indices. Table 1 lists a few of the indices we developed (see McNamara et al., 2014 for more information).

In some respects, our goal was essentially realized when we conducted a Principal Component Analysis that included the indices we had tested during these years of analyses (Graesser et al., 2011). This analysis revealed two major findings: 53 Coh-Metrix indices accounted for 67.3% of the variance in the large corpus of texts (i.e., Touchstone Applied Science Associates, TASA) and the indices loaded onto five principal components that were well aligned with theories of text and discourse (i.e., narrativity, referential cohesion, syntax, word concreteness, and deep cohesion). The former shows that a very large amount of variance (i.e., differences and similarities) between texts is related to factors associated with text difficulty (or ease), whereas the latter confirmed the success of our approach. The first component, narrativity (accounting for 18.5% of the variance), included indices related to differences between genres such as narratives and informational text. The second component was referential cohesion, and it accounted for 14.1% of the variance: more than syntax or word concreteness. The importance of text genre (i.e., narrativity) is well established. Beyond genre, our analyses and experimental studies have consistently demonstrated the additional and orthogonal importance of text cohesion.

The orthogonal contributions of genre and cohesion are important. One might assume that texts with greater narrativity are more cohesive; they are certainly easier to understand. In fact, informational and narrative texts are equally likely to vary in cohesion. However, narrative texts tend to be *lower* in cohesion than science texts (McNamara et al., 2012, 2014). This makes sense when readers' knowledge is considered. When readers' have more knowledge about the information in the text, cohesion gaps induce the reader to infer the connections between the ideas. For narratives, these inferences can help to make the passage or story more interesting. Inferences also help to improve memory for the text and in turn improve learning. For narratives, sufficient knowledge to bridge cohesion gaps can be safely assumed. Writers naturally tend to include conceptual gaps in text when they know readers are familiar with the domain (McNamara, 2013). In contrast, cohesion is more important for science texts. Indeed, over 20% of the variance between science texts is accounted for by the presence or absence of referential cohesion. Cohesion also tends to be higher because it is a key to comprehension and learning, particularly for low knowledge readers who seek to learn the information.

The development of Coh-Metrix afforded researchers and educators a tool that provides measures of and in turn means to adjust text cohesion (Figure 1). TERA (Text Easability and Readability Assessor; http://www.commoncoretera.com/) provides researchers and educators with a profile analysis of text ease based on the Coh-Metrix component scores. Coh-Metrix has also inspired researchers to develop their own natural language processing tools. For example, Kristopher Kyle and Scott Crossley developed the Tool for the Automatic Analysis of Text Cohesion (Crossley et al., 2016b) that provides about 150 indices related to text cohesion, including indices related to lexical and semantic overlap, connectives, and lexical diversity, at both local and global levels. Such tools provide

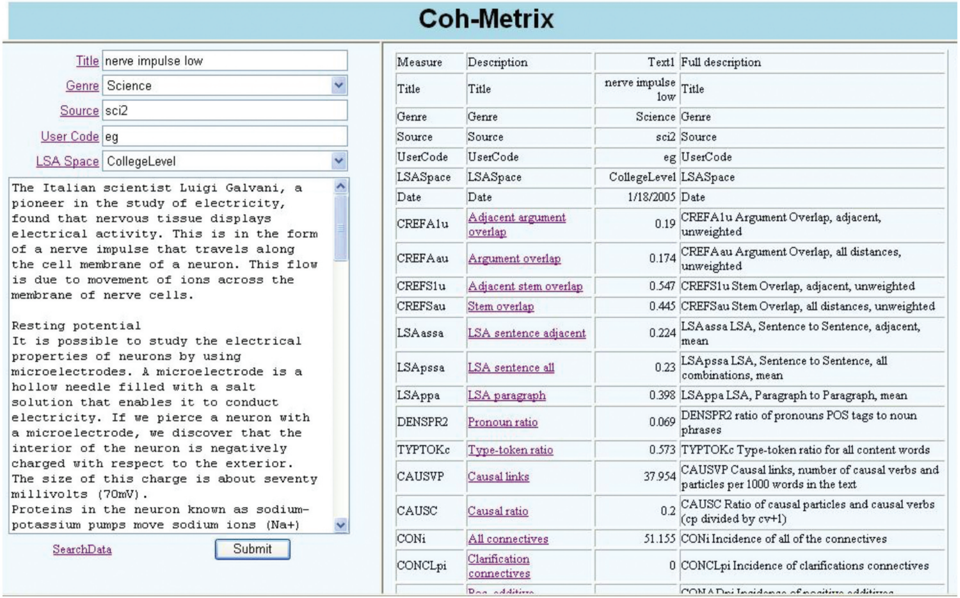

Figure 1. Coh-Metrix allows researchers to investigate multiple aspects of text and discourse by extracting features of language related to ease of comprehension.

the means to explore the effects of cohesion and many other facets of text at multiple levels and across multiple contexts.

The reader

After it leaves the writer, text does not exist without a reader, and no text is the same for any two readers. We can describe text objectively, using terms such as high or low cohesion, simple or complex syntax, familiar or rare words. However, these constructs and measures are only meaningful relative to what the reader brings to the table. Hence, individual differences are a crucial consideration when investigating comprehension.

On the surface it may seem that knowledge of and ability to remember words in text must be the most important variables to consider. Indeed, research on the ability to decode, attend to, and remember words leads many researchers to believe that the important individual differences of interest are constructs related to attention and working memory (McNamara, 2020; McNamara & O'Reilly, 2009). If the task requires remembering words, then the difficulty of the words (e.g., word frequency) and the reader's vocabulary knowledge will account for a large amount of variance. However, words in sentences are magic—once you put words in sentences, and then in multiple sentences, and string them together to provide coherent meanings, the effects of lower-level cognitive processes fade away (Healy et al., 1987) and variance in other processes related to comprehension take over the stage (e.g., Allen et al., 2014b).

Readers vary in many ways. The research that I have conducted to investigate these differences, in particular how to overcome the challenges that readers experience, was inspired by two theoretical camps. First, of course, there was the CI model. The CI model principally emphasizes the effects that emerge from differences in prior knowledge. When readers have less knowledge, the sparsity of the situation model increases, leading to lower comprehension, particularly in the absence of text cohesion. The second inspiration came from theories emanating from Educational Psychology researchers such Alexander, Brown, Scardamelia, Bereiter, and Palincsar. Of these, Palincsar and Brown (e.g., Palincsar & Brown, 1984) brought comprehension strategies to the forefront. In turn, Pat Alexander was among the few in the early 1990s who stressed the importance of domain knowledge (e.g., Alexander et al., 1994). These researchers were in a different world from my own, because at that time, I lived in the world of Cognitive Science. However, their work was music to my ears. It helped me to conceive of Self-Explanation Reading Training and in turn its automated version, iSTART.

Technology: iSTART

Interactive Strategy Training for Active Reading and Thinking was developed based on Self-Explanation Reading Training (SERT), which comprised instruction and practice on how to use comprehension strategies while reading science texts (Figure 2). This intervention was inspired primarily by the work of Palincsar and Brown (1984) but was enabled by the work on self-explanation. I had first heard of self-explanation in the early 1990s when Michelene Chi gave a talk on the topic at the University of Colorado. However, the true inspiration came primarily from a study by Kate Bielaczyc et al. (1995), who demonstrated the benefits of combining self-explanation with self-regulation strategies on improving students' computer programming skills.

On the one hand, there was abundant evidence that comprehension strategy training improved reading comprehension, but the training required a large number of classroom practice sessions (Palincsar & Brown, 1984). On the other hand, self-explanation effectively improved problem solving and learning, but only for the elite few who did it well (e.g., Chi et al., 1994). In turn, SERT was designed like a Reese's Cup: combining comprehension strategies and self-explanation like peanut butter and chocolate, symbiotically. Comprehension strategies improve less skilled students' ability to explain text to themselves, and self-explanation externalizes the strategies, making them visible and repairable while at the same time encouraging causal processing. It was long shot, but the McDonnell foundation provided the initial funding to explore the possible benefits of combining the two. It worked. SERT has been shown to improve comprehension of challenging texts, learning from texts, and performance in science classrooms, with relatively few training and practice sessions (McNamara, 2004, 2017). Moreover, SERT primarily benefits low-knowledge and less skilled readers (Magliano et al., 2005; McNamara, 2017; McNamara et al., 2006). These results demonstrate that generating self-explanations is effective, but readers need to also learn comprehension strategies to capitalize on the benefits.

SERT worked, but it was not scalable. Implementing the instruction on a large scale required a level of patience and expertise that I just did not have. The experimenters (i.e., instructors) had to know how to provide the instruction (and show up), the students just had to show up, and the TV had to be there and turned on (some demonstrations were provided in videos on eight-track). For example, in one of our intervention studies, we arrived at the remote, rural school in Kentucky (after an 8-hour drive) and they informed us that school had just qualified to play for the state championship in football at the end of that week. Yay! We had to reduce the study down from 5 full days to 4 half days, and needless to say,

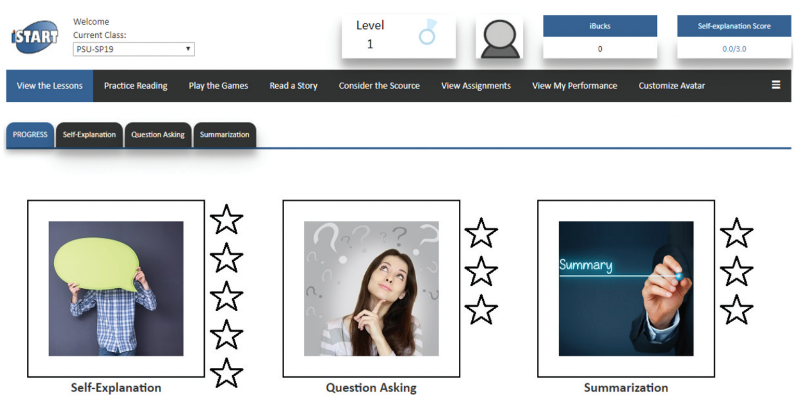

Figure 2. Interactive Strategy Training for Active Reading (iSTART) provides instruction and practice on how to use comprehension strategies while reading challenging texts.

many students were absent or distracted. Despite the scheduling nightmares and distractions, SERT was still shown to be beneficial (McNamara, 2004), but the stress exceeded my capacity.

I was inspired by the success of SERT and the complications in its delivery to develop the automated version of SERT, which I called Interactive Strategy Training for Active Reading and Thinking, or iSTART (Levinstein et al., 2007; McNamara et al., 2004). I gave it this name because I envisioned an automated, intelligent tutoring system that went beyond self-explanation and comprehension strategies and also provided instruction and practice to improve students' ability to evaluate information, and think critically (McNamara et al., 2006, 2007b).

iSTART provides instruction for five comprehension strategies in the context of self-explanation: comprehension monitoring, paraphrasing, prediction, bridging, and elaboration. Comprehension monitoring is the readers' ability to assess their understanding of the text while reading. Paraphrasing is a restatement of the text in a reader's own words. Prediction is when a reader anticipates forthcoming information in a text either by making educated guesses or taking note of information that, if present, will aid in comprehension of a previous concept. Bridging is the act of drawing connections between the current sentence to previous information in the text. Elaboration is using prior knowledge, either general or domain-specific, or logic to expand on the concepts in the text. iSTART first instructs students on the five comprehension strategies using video lessons that provide students with information about comprehension strategies to prepare them to practice the strategies in both regular (*coached*) and game-based practice.

iSTART includes two types of game-based practice to increase motivation and engagement (Jackson & McNamara, 2013) (Figure 3). In generative games, students earn points for producing high quality self-explanations. In identification games, students read example self-explanations of a text and earn points by correctly identifying which reading strategies were used in the examples. To further increase agency, students can use their points to unlock new games and purchase customization features for students' avatars. These "metagame" elements were designed to promote student motivation (Jackson & McNamara, 2013).

During regular practice, students read a text and type self-explanations for certain target sentences. The self-explanations are analyzed using natural language processing (NLP) algorithms that use both word-based indices and latent semantic analysis to identify the strategies used in the self-explanation. The algorithm provides a holistic score for the quality of the self-explanation on a scale of 0 to 3 (0 = poor, 3 = great) as well as feedback messages. If the student's self-explanation score is less than 2, the system will also select and provide actionable feedback and allow students to revise their self-explanation (Boonthum et al., 2007; McNamara et al., 2007a).

iSTART can also adapt the difficulty of texts based on students' performance in iSTART. For example, when students' self-explanation quality is high, the subsequent text will be more difficult. Conversely, when students' self-explanation quality is low, the subsequent text will be less difficult (Balyan et al., 2020; Johnson et al., 2018). The addition of adaptive, just-in-time support leads to an increased sense of learning (Watanabe et al., 2019) and demonstrates positive learning outcomes, particularly for less skilled readers (McCarthy et al., 2020b).

In addition to these self-explanation strategies, iSTART has recently been expanded to include instruction and practice for two additional macro-strategies: question-asking and summarization (Johnson et al., 2017) with automated feedback (Crossley et al., 2019; Ruseti et al., 2018b, 2018a). Question asking serves as a starting point for students to generate inferences (McCrudden & McNamara, 2017). Having readers develop questions about the text encourages comprehension and induces linking of ideas across sentences (Rosenshine et al., 1996). Students who ask higher quality questions recall more information from the text and answer more questions correctly about the text. Teaching students, particularly developing readers, the importance of asking questions, how to ask questions and answer their questions, and how to evaluate the quality of their questions leads to improved comprehension and metacomprehension accuracy.

While question asking and explanation encourage inferencing as a means of elaborating the text with more information, the goal of summarization is to reduce the text to its core ideas. This process helps

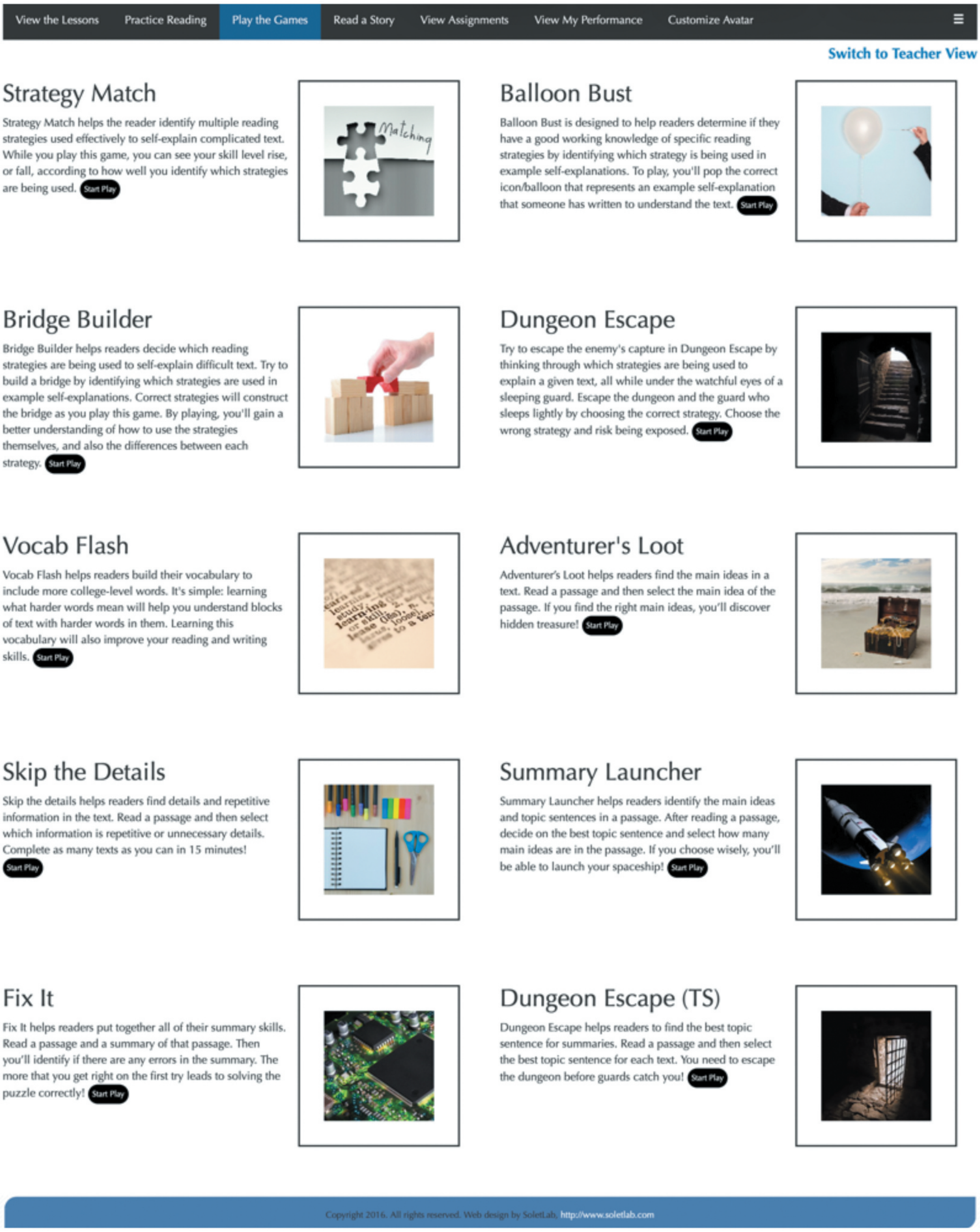

Figure 3. Game-based practice in iSTART includes games that provide opportunities to identify effective comprehension strategies or practice generating constructed responses (e.g., explanations, summaries) while reading challenging texts.

readers identify irrelevant information, integrate content with preexisting knowledge, and better retain text material. Furthermore, summarizing reinforces readers' mental representations of the content, enhancing not only retention of text material but also conceptual understanding, particularly for lower achieving students and those with learning disabilities. These benefits are summarized by Graham and Hebert (2011), who reported that summarization enhanced comprehension in 18 of 19 studies in their meta-analysis.

iSTART has been shown to improve self-explanation quality and reading comprehension for readers from middle school through adulthood and is particularly beneficial for low-knowledge and less skilled readers (e.g., McCarthy et al., 2018, 2020b; McNamara et al., 2007b; Snow et al., 2016). The focus of a current project in collaboration with Panayiota Kendeou and Carol Connor[1] is to develop iSTART for developing readers from grades 3 to 5. iSTART-Early provides game-based comprehension strategy instruction and practice, including question asking, paraphrasing, explanation, and summarization (Figure 4). Our long-term objective with iSTART is to cover the developmental spectrum of readers from young developing readers (i.e., iSTART-Early;Institute of Education Sciences R305A190050) to adult literacy learners (iSTART-ALL; Office of Naval Research N00014-17-1-2300; N00014-20-1-2623). Across the developmental spectrum, the focus of iSTART remains on comprehension strategies, with the theoretical assumptions that a key ingredient of comprehension is the ability to make inferences (McNamara, 2020) and strategies are key to learning how to deeply comprehend text (McNamara, 2009).

Theory. Multiple theories inspire and drive iSTART. Similar to SERT, it was inspired by theories of text and discourse (principally the CI model) and by theories undergirding comprehension strategy work (principally Knowledge Building theories; Scardamalia & Bereiter, 2006). Most relevant to iSTART is the notion that "idea improvement" is an explicit principle and objective rather than something that remains implicit in learning tasks or activities. Vygotskian notions of externalizing, sharing, and building on ideas were central to iSTART as well.

How can you externalize, share, and build on ideas in an online tutoring system (particularly in 2001)? Interestingly, the notion of implementing assumptions of knowledge building in an automated system can seem antithetical to theories of knowledge building. In the same way, the notion of *self-*

Figure 4. iSTART-Early provides comprehension strategy instruction environment where students travel through space to get back to Earth. Each planet provides comprehension strategy instruction and game-based practice, including question asking, explanation, and summarization.

explanation is in some sense antithetical to a *community* of knowledge builders. However, my assumption was, and remains, that individuals cannot contribute to a community of knowledge builders if they are not given the tools to contribute, and I assume that the ability to understand challenging text is a necessary tool to bring to the table.

One challenge was how to provide meaningful feedback to students regarding the quality of their responses with regard to their use of knowledge building strategies without challenging the accuracy of the content. For me, the answer at the time lay in Latent Semantic Analysis (LSA), which provided the ability to represent semantic relations between ideas in text by transforming text into vectors and extracting the dimensions using a mathematical equation called singular value decomposition. I was inspired by Art Graesser's success in the use of LSA in AutoTutor (Graesser et al., 1999). LSA is very powerful, but in the end it fell far short of providing the complete solution. Providing feedback to natural language within a tutoring system is far more complex than simply deriving estimates of semantic overlap. It requires careful, iterative considerations of multiple aspects of students' responses in combination with considerations of the student model, pedagogy, and pedagogical goals. McNamara et al. (2007a) describes the history of our attempts and our initial algorithms. We combined both LSA and word-based algorithms using machine learning (i.e., discriminant function analysis) and implemented a relatively complex set of feedback algorithms that interactively prompt the readers to improve their self-explanations (Millis et al., 2007). The current feedback algorithms are guided by theories of skill acquisition and pedagogy and many iterations of user testing.

The other side of iSTART, its face, is the tutoring system. My approach to implementing tutoring within iSTART was principally guided by learning principles derived from research in memory and skill acquisition, which goes back to my graduate training with Alice Healy, Lyle Bourne, and Anders Ericsson. In tribute to my graduate advisor, Alice Healy, McNamara et al. (2015b) describe three of these principles in their relation to iSTART (i.e., the generation effect, deliberate practice and feedback, and antidotes to disengagement). An underlying theoretical premise is that skills (and strategies) cannot be learned without deliberate, repeated (but spaced), generative practice with formative feedback (Healy et al., 1993). Accordingly, a key component of deliberate practice is individualized, targeted, and actionable feedback that informs learners with information about what needs to be improved and how to improve it (Ericsson et al., 1993). In iSTART, students are provided with feedback and encouragement during practice by assessing the degree to which their responses meet certain objectives as well as considerations of the individual student's learning path.

Unfortunately, some learning paths can be long and tedious. Practice can, over time, become tedious for students; thus, deliberate practice requires *antidotes to disengagement* (Healy et al., 2012). For the latter, iSTART incorporated game-based practice. Games can be inherently designed to increase players' depth of cognitive engagement, and they are excellent platforms for incorporating principles that enhance learning (e.g., random, spaced testing; implicit and explicit feedback; part-task and whole-task practice; leveling; contextualization and self-directed objectives). My quest to incorporate games into iSTART began with Tanner Jackson and Art Graesser (McNamara et al., 2010b). Our initial proposal was to build an immersive game using a narrative structure. We wanted to build a game that incorporated the principles of iSTART but within one narrative. While the proposal was awarded by the National Science Foundation, we were only provided with half the funding requested. The National Science Foundation asked us to reconsider our objectives with the resources provided; in response, we proposed building short dynamic games that provide students with practice on the comprehension strategies. In the end, this better matched the structure of iSTART and allowed us to build games that varied game elements and conduct experiments to examine the impact of game features (Jackson et al., 2015; Jackson & McNamara, 2017; McNamara et al., 2015b; Proske et al., 2014). We found that the particular features had little effect on instructional gains, but that games combined with opportunities to personalize system features enhance students' engagement and motivation, affording a sense of agency and personal investment in their learning progress (Jackson et al., 2015; Jackson & McNamara, 2013; McNamara, 2017).

In summary, the core of iSTART was inspired by theories emanating from multiple fields including text and discourse, education, linguistics, and cognitive psychology. Building a tutoring system for such a complex process as comprehension requires a pluralistic approach. iSTART is theory driven, but no one theory or domain would have sufficed.

The writer

Next, we turn to the writer. I've drawn the analogy between the writer and a diamond cutter. Diamond cutting is the practice of shaping a diamond from a rough stone into a faceted gem. Like writing, it requires highly specialized knowledge and techniques due to its extreme difficulty. My work developing Coh-Metrix and iSTART led me to think about the writer. I assumed at first (somewhat incorrectly) that indices such as cohesion in Coh-Metrix would be key to analyzing writing and, second (correctly), that students need access to instruction on strategies to produce high quality text and practice with formative feedback.

My initial interest in writing and motivation that eventually led to the Writing Pal was inspired by conversations with Ron Kellogg. Together, we resubmitted a proposal to IES two times. In this version, Ron and I were simply proposing to use iSTART to investigate its effect on the writing. This idea (fortunately) did not sell. Subsequently, I worked with Phil McCarthy (who understood writing better than I did) to write the proposal to develop the Writing Pal, wherein the strategy training was focused on writing processes. We submitted this version three times to IES, and it was finally awarded in 2008. In sum, the Writing Pal was born from collaboration, persistence, and listening to the proposal reviewers.

When we began conceiving of the Writing Pal, little existed to inform its development. Writing is a crucial part of our lives and is undeniably considered to be a *discourse process*, yet the first edition of the *Handbook of Discourse Processes* in 2003 featured no chapters on the topic of writing, and, when mentioned, it was only tangentially as a means of communication. It was not a focus of research for any of my colleagues in text and discourse, cognitive science, or cognitive psychology: in sum, as discourse scientists, we were in new territory.

The first edition of the *Handbook of Writing Research* (MacArthur et al., 2006) was published just as we submitted our third version of the proposal, but there was little research to inform the development of an automated, intelligent tutoring system for writing. There were two (very) separate camps of writing researchers (Graham, 2018; McNamara & Allen, 2017). One camp advocated primarily from *sociocultural* perspectives, emphasizing social factors such as audience, purpose, and medium. The other camp considered the *cognitive* perspective of writing, with a large emphasis placed on the interplay of memory, retrieval, and processing, largely from an information storage perspective (which I had long ago abandoned; cf. Healy & McNamara, 1996). Much of the research on the use of strategies was from the perspective of special education (Graham & Harris, 2003). The latter research combined with consultations with Steve Graham helped to guide development of the Writing Pal.

Technology: the Writing Pal

The Writing Pal provides writing strategy instruction along with deliberate practice for high school and early college students. Relative to other automated systems for writing (see Allen et al., 2016b, for a review), the Writing Pal is unique in its focus on explicit strategy instruction and its varied opportunities for practice (i.e., game-based strategy practice and essay writing practice). Strategy instruction is delivered via lesson videos on the principle phases of the writing process: prewriting, drafting, and revising (Roscoe et al., 2014, 2015) (Figure 5). These videos explain and demonstrate a variety of principles and strategies (see adaptiveliteracy.com). The Writing Pal *works*. When students engage with the Writing Pal, we have observed increases in strategy knowledge, use of revising strategies, and essay scores (e.g., Crossley et al., 2016a; Roscoe & McNamara, 2013; Roscoe et al., 2015b).

By the time we began developing the Writing Pal, we had already begun the process of developing iSTART games and were already convinced that games were an ideal means of motivating students during practice in conjunction with well-informed pedagogy. Thus, we developed a suite of strategy practice mini-games (Roscoe et al., 2013a) (Figure 6). The purpose of these games is to reinforce

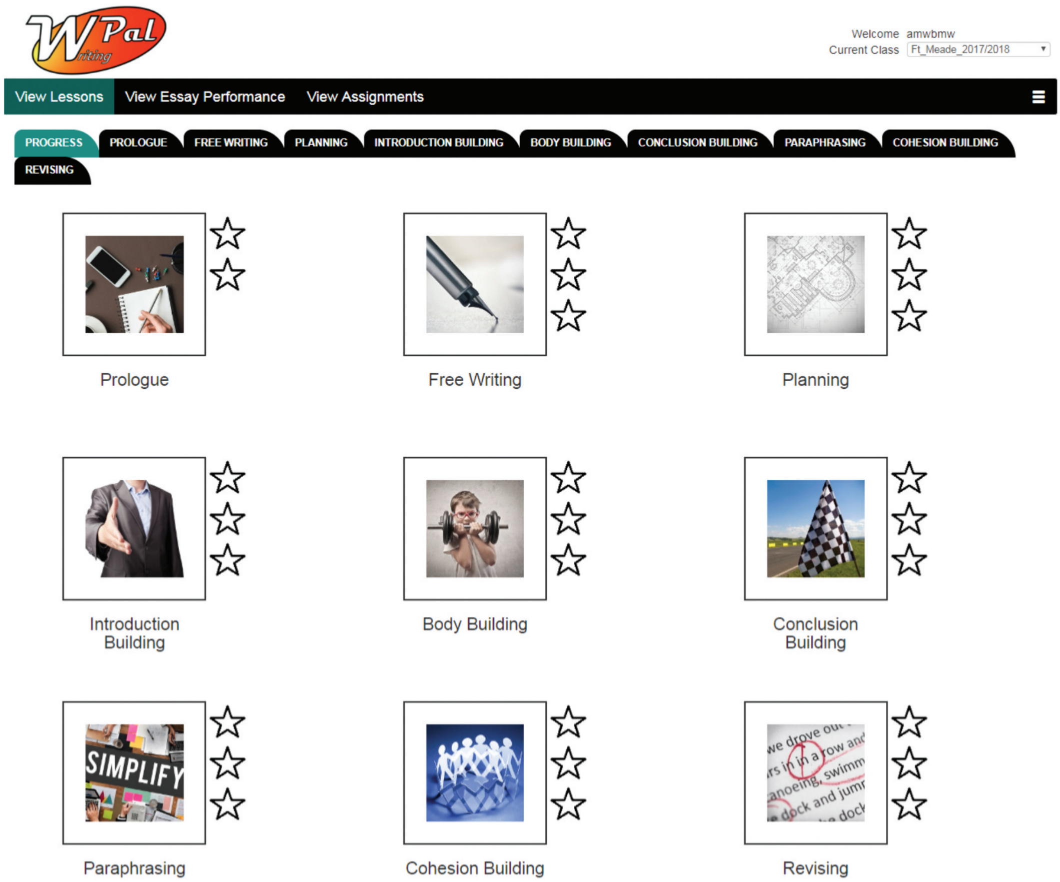

Figure 5. The Writing Pal provides nine modules that include instruction on strategies and game-based practice related to the three phases of writing: prewriting, drafting, and revising.

writing strategy knowledge through games wherein students identify examples of good and poor writing or the strategy(-ies) used to improve writing (Allen et al., 2014a; Proske et al., 2014; Roscoe et al., 2019). Students also have the opportunity to practice writing persuasive essays along with automated summative and formative feedback.

When we conceived of the Writing Pal, we were focused on providing strategy instruction. However, it was quickly evident that writing practice with formative feedback is key to learning how to write (see Figure 7). Developing automated writing evaluation (AWE) algorithms emerged as a major focus within the project. AWE involves using NLP to extract linguistic features from the essays, and series of algorithms are used to assess overall quality and in turn guide feedback (McNamara et al., 2015). When we began development of the Writing Pal, we only had Coh-Metrix indices at our disposal to use within our AWE algorithms. However, Coh-Metrix was intended to assess text difficulty, not quality, and so it fell quite short of explaining a sufficient amount of variance to provide feedback on writing quality (McNamara et al., 2010a). As such, Scott Crossley, Rod Roscoe, and I focused on developing indices related to writing (rather than text difficulty) and testing those indices' contributions to NLP algorithms designed to provide feedback to developing writers.

Over the past decade, we have developed numerous new indices and new NLP tools (e.g., Crossley et al., 2015; Crossley & McNamara, 2014; Crossley et al., 2014; McNamara et al., 2013, 2015). These NLP tools have contributed to the development of several iterations of summative and formative algorithms that drive feedback during essay practice within the Writing Pal. Summative feedback is provided as

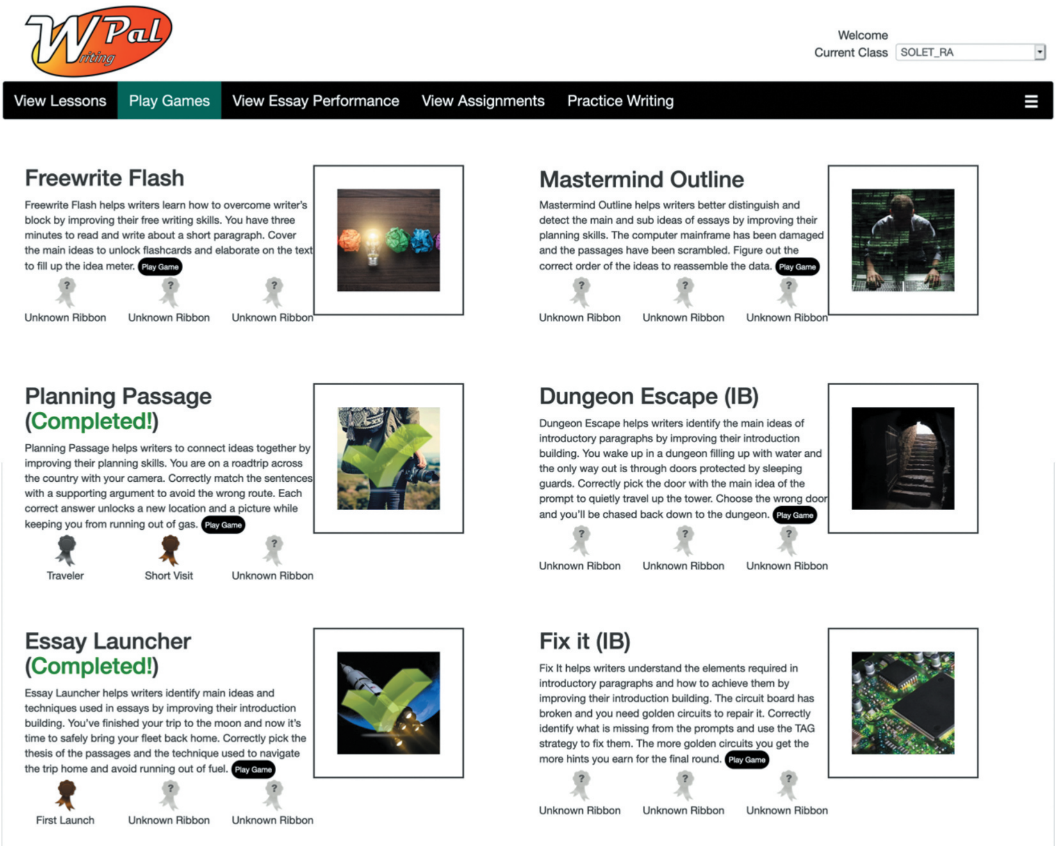

Figure 6. Game-based practice in the Writing Pal offers students opportunities to practice writing strategies within a variety of games. Over 20 games are offered in the Writing Pal aligned to modules that cover strategies to plan, draft, and revise essays.

a holistic score on a 1 to 6 scale, and formative feedback is given at the essay-level (e.g., length, relevance, structure) and section-level (e.g., introduction, conclusion). The feedback is designed to be specific, actionable, and, most importantly, aligned to strategies taught in the lessons (e.g., Roscoe et al., 2013b).

Theory. The fundamental theory of change that drove the development of the Writing Pal is that learning writing strategies is key to increasing the likelihood that less skilled writers' will catch up to skilled writers. Strategies provide shortcuts, circumventing the need for hundreds of hours of practice with feedback from devoted instructors that would be necessary to catch up, which is often absent from many students' educational experiences. Hence, the Writing Pal includes instruction and game-based practice on writing strategies and feedback that refers back to those strategies. A key aspect of the Writing Pal is that the feedback students receive always gives them specific strategies they can use to improve their writing, not solely what is wrong with their essay.

The Writing Pal was also guided by pedagogical practices in writing (literally, by scouring writing textbooks) and theoretical assumptions regarding writing processes. According to most theories of writing, it comprises three main stages (i.e., planning, drafting, and revising) that are intertwined and often co-occur. Writing is among the most complex cognitive and social processes, combining knowledge of language and the world, reasoning, decision-making, and, importantly, how to write. Many or most developing writers have little knowledge on how to approach this complex process. Our objective was to break this down, and provide writers with strategies on how to accomplish each of these writing subgoals (e.g., planning, drafting, and revising). In line with theories of skill acquisition, this approach

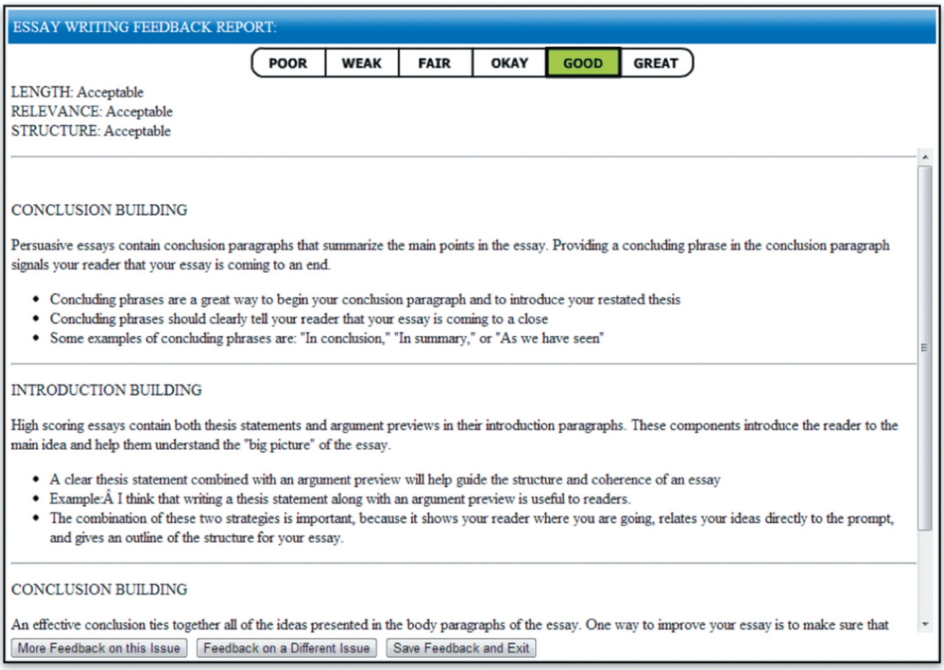

Figure 7. The Writing-Pal offers students opportunities to practice writing essays with summative and formative feedback on strategies that they can use to improve the essay and their writing skills.

implements the use of part-task training (i.e., subgoals and strategies) combined with whole task practice (i.e., writing essays). Accordingly, the Writing Pal consists of strategy training modules corresponding to each of the subgoals. Each module culminates with writing an essay for which the feedback centers on how the student can use the Writing Pal strategies to improve the essay (or subsequent essays).

One objective of providing strategy training is to guide students in tackling complex, seemingly overwhelming tasks. Another is to provide multiple paths to completing the task: flexibility in having multiple routes to success. Skilled writers are able to adapt their writing to the intended audience (McNamara, 2013). Skilled writers also adapt their writing to their own knowledge and skills relative to the task; they are flexible. There is a prevailing assumption that high quality writing has identifiable markers, which in turn translates to either reliable scores by experts or an automated algorithm to score essays. While this is demonstrably true, skilled writers can use multiple techniques to construct a higher quality essay. For example, in Crossley et al. (2014), we found four clusters of higher quality essays based on their linguistic features. One subset of writers adopted a more academic style, with complex syntax and more unfamiliar, academic words—writing characteristic of texts that are more challenging to comprehend. A second cluster was characterized by the use of a large number and diversity of unfamiliar words but with a strong semantic overlap with the essay prompt and word choices typical of higher quality essays, essentially showing off their sophisticated lexical choices. A third subset of writers used more accessible language with high cohesion: writing characteristic of texts that are easier to comprehend. A fourth approach was more narrative in style, with action and imagery, lacking cohesive cues but facilitating comprehension and engagement with a more story-like structure.

Skilled writers more flexibly change strategies according to the demands of the essay topic. Allen et al. (2016d) further demonstrated that more skilled writers flexibly used narrativity within their essays across eight prompts (see also Allen et al., 2019). That is, sometimes more skilled writers used narrativity within the essays and sometimes did not. Some less skilled writers use narrativity and some do not, but they less flexibly vary narrativity across essays. Additionally, Allen et al. (2016d) found that less skilled writers increased in flexible use of writing features in their essays as a function of strategy training within the Writing Pal. Providing strategy training helps students to learn multiple ways to write an essay, such that they can learn to adeptly assemble their available knowledge and skills unique to each prompt and task. Whereas using a story-like structure may be more useful in some cases, relying on domain knowledge with specific evidence may be feasible in others. Skilled writers have learned how to pull resources together depending on the demands of the task at hand.

When we began my quest to understand writing processes, I was quite frankly naïve. For example, I assumed that cohesion would play a critical role in explaining text quality. This expectation was largely driven by my analyses of published texts: Cohesion explains a substantial amount of variance between published texts and has substantial impacts on comprehension (Graesser et al., 2011; McNamara & Kintsch, 1996). It was also driven by both pedagogical and theoretical literature regarding writing, which emphasized cohesive writing. One of the biggest lessons that I learned in analyzing writing has been that text difficulty and text quality are orthogonal constructs; more challenging text (e.g., low cohesion, rare words, syntactically complex) is often judged to be of higher quality than easy-to-read text. Cohesion plays a vastly different role in text quality in contrast to the role it plays in text difficulty. I was encouraged by Teun van Dijk (at his conference in Barcelona) to consider these differences in terms of epistemological theories (McNamara, 2013). Within such a conceptual framework, the epistemic frame, goals, and skills of the author drive the knowledge demands and quality of writing. Skilled writers are aware of their audience and adjust the demands of the text according to the knowledge and skills of the audience. Consequently, when the target audience has more knowledge of the contents of the text (e.g., for stories and narratives), text generally comprises more challenges in terms of cohesion and syntax. When the content has greater knowledge demands (e.g., science), writers compensate for those demands by using more cohesive cues (e.g., connectives, semantic overlap) and simpler syntax.

Multiple speakers, multiple texts

The bulk of my work has considered individual readers reading one text or individual writers. My current work attempts to understand the intersections of the reader and the writer (or speaker) and cases when there are multiple texts or speakers (or a *diamond mine*). Considering the intersections of the reader and the writer brings to the forefront multiple issues including the relations between reading and writing, the role of communication in writing and dialogs, comprehension of multiple documents and social media, and source-based writing skills and strategies. Below I describe the technologies associated with these objectives.

Technology: cohesion network analysis

The intersection of the writer and the reader brings to the forefront the role of communication in writing and dialogue. My work on communication, collaboration, and dialogue was largely spurred by Mihai Dascalu and his mentors Philippe Dessus and Stefan Trausan-Matu. I met Mihai around 2013 at AIED (Artificial Intelligence in Education) where he showed me his initial work toward developing *ReaderBench*, a tool to provide multilingual linguistic analytic tools, primarily centered around the importance of cohesion. Fortunately (for me), Mihai was subsequently awarded a Fulbright to spend a year working with me at Arizona State University, where we began our collaboration on the development and testing of a suite of tools related to language, and in particular Cohesion Network Analysis (CNA).

CNA combines NLP with Social Network Analysis (SNA) to analyze discourse structures using measures of text cohesion (Dascalu et al., 2015; Dascalu et al., 2020a). SNA represents and examines social structures using graph theories. Similar to the CI model, SNA characterizes networks in terms of nodes (individual actors, people, or things within the network) and the edges or links (i.e., interactions, relationships) that connect them. CNA defines those links in terms of cohesion and uses SNA-derived

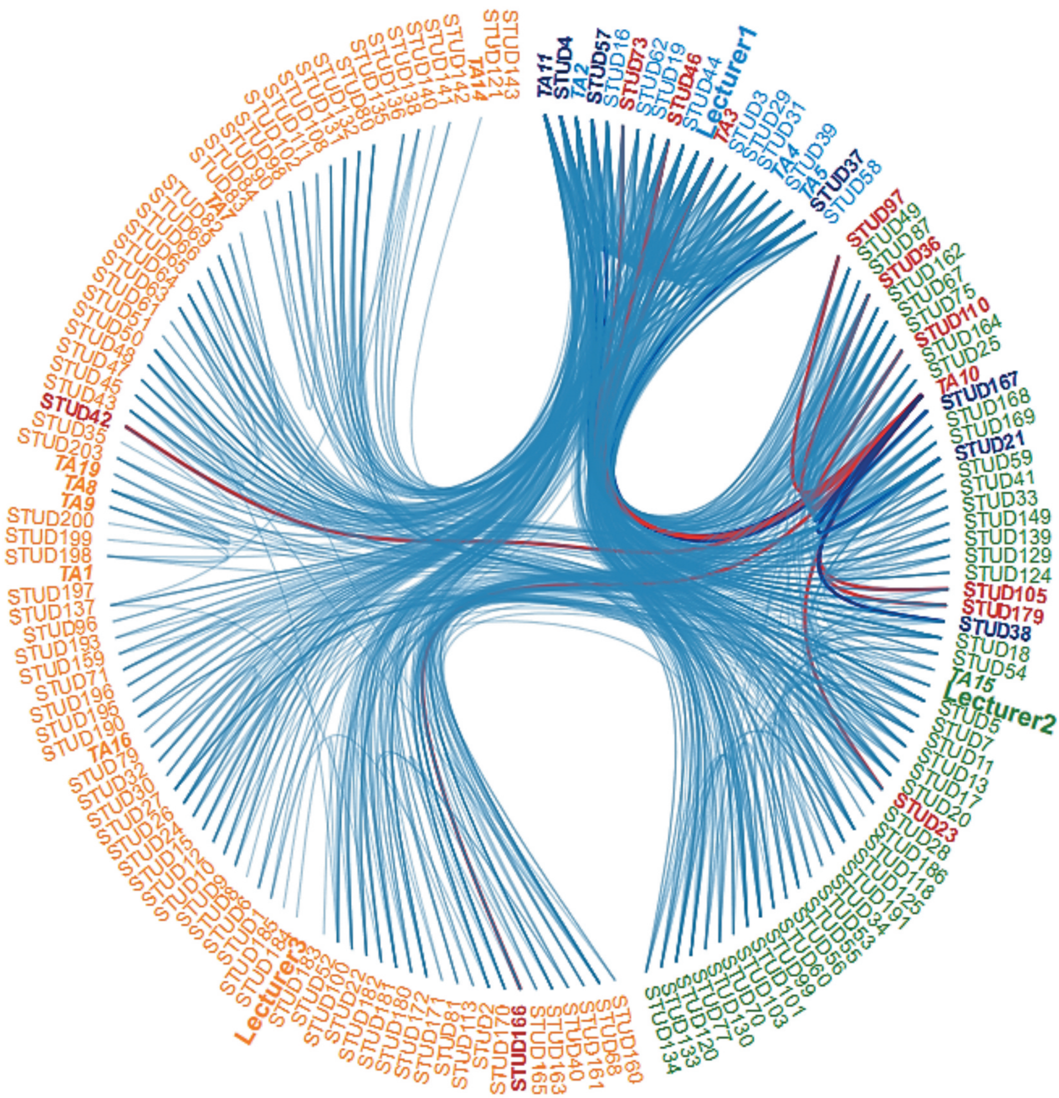

Figure 8. Cohesion network analysis graph. CNA defines connections between a network of individuals in a network in terms of cohesion. This graph illustrates the semantic connections on a discussion board between students, teaching assistants, and lecturers in an online course.

indices to describe the network of relationships and the flow of information within a network of individuals (Figure 8). Cohesion is computed using various similarity measures from different semantic models such as Latent Semantic Analysis, Latent Dirichlet Allocation, and word2vec. The cohesion graph is a multilayered structure wherein a central node represents the conversation's thread comprising multiple contributions, which are further divided into sentences and words. Various computations can be used to denote the relevance of a contribution in a conversation or the impact of a word within a sentence or contribution.

We have used CNA in various studies of collaborative dialog in the context of online courses' discussion boards (e.g., Dascalu et al., 2018b; Dascalu et al., in press), blended math courses (Crossley et al., 2018), and between researchers (Paraschiv et al., 2017). We have also used CNA to model semantic overlap between ideas within texts (Dascalu et al., 2018a, Dascalu et al., 2020b) (Figure 9) and between texts (Dascalu et al., 2020b; Nicula et al., 2019, 2020) (Figure 10) and extract the main ideas from texts

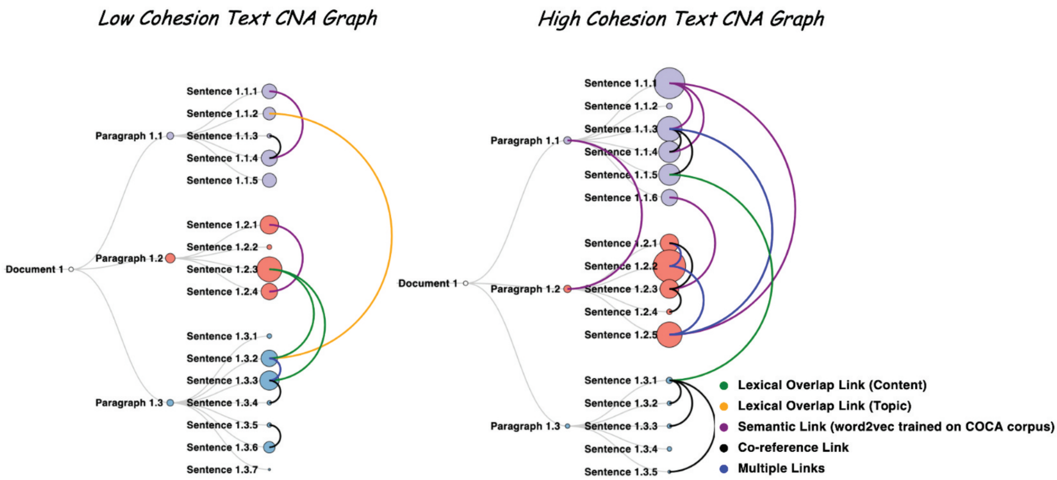

Figure 9. Cohesion network graph of low and high cohesion texts. CNA calculates the connections between ideas in texts in terms of multiple levels, including lexical, semantic, and co-referential links between ideas. This graph illustrates the differences between high and low cohesion texts on the same topic (see M.-D. Dascalu et al., 2020b).

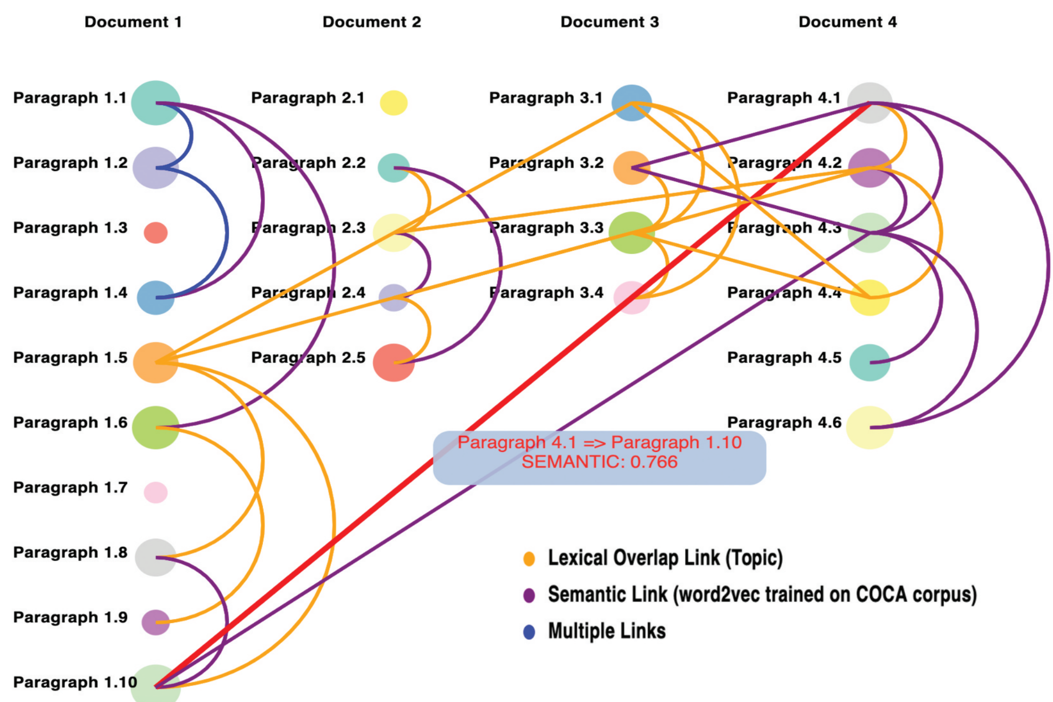

Figure 10. Multidocument cohesion network analysis graph. MD CNA calculates lexical and semantic links both within and between documents.

(Cioaca et al., 2020). In sum, CNA is a powerful method of representing the connections between ideas and the flow of a dialog, essay, or text.

Theory. CNA is similar to the CI model in that the model describes a network of ideas and the links between them. Within the CI model, the nodes are the nouns, arguments, and propositions and the links between them are defined by the *prepositions* [e.g., *in*(David, office)], verbs or actions [e.g., *accept*(David, paper)], or states [e.g., *is*(Danielle, happy)] and spreading activation to semantically related concepts expands the network to ideas that are not explicit within the text or discourse (depending on the reader and the context). However, within CNA, the links are defined solely by semantic overlap, and thus verbs (i.e., actions) have a comparable role within the network as do nouns (i.e., things, people). The syntax that glues the words together is represented as well within a multidimensional network.

While there were similarities between CNA and my previous work modeling comprehension based on the CI model and Coh-Metrix, my work with Mihai Dascalu and his colleagues introduced me to a different way of thinking about text, cohesion, and communication in light of Dialogism (Bakhtin, 1981). On the surface, CNA can be simply described as a representation of links based on cohesion. However, it is not so simple. I spent many hours wrapping my head around the notion of voices and the intertwining synergies of different *speakers* in text. On the surface, dialogism should only be relevant to research on dialogs: interactions between multiple speakers. It is indeed relevant to research on collaboration, online chats, discussion boards, and so on; however, it is also relevant to understanding the flow of information within text and between texts. My work on cohesion had already occurred within the theoretical framework of the CI model, which is based on a network of information and the flow of activation within that network. However, discourse modeled from a dialogical perspective further considers interactions as building meaning and understanding within a network composed of multiple layers and dimensions.

In terms of dialogism, the main goal of a discussion can be described in terms of voice interanimation and *polyphony* in which conflicting views, various angles, and multiple perspectives co-occur; all the previous aspects should also be covered in a truly collaborative conversation. However, as voices express ideas and opinions, polyphony can be used to perform a deep dialogical discourse analysis by summing up multiple voices co-occurring within the same discussion thread. A longitudinal dimension is reflected in the explicit or implicit references between utterances, following the conversation timeline. This grants an overall image of the degree of interanimation of voices spanning the discourse. Thus, polyphony can be used as a signature for collaboration, as the interactions between multiple participants of the conversation are reflected in their *voices*. The notion of discussion threads enables highlighting the evolution of voices across time. Finally, a transversal dimension is used to represent a differential positioning of participants or ideas, when a shift of points of interest or views occurs and there is a move toward discussing other topics. In sum, a discussion or text cannot be modeled as one idea at a time in sequence. It is the interanimation of multiple ideas and voices that co-occur, simultaneously, in harmony or out of harmony—which is polyphony.

Technology: writing assessment tool

Considering the intersection of the writer and the reader also led me to interests in multiple document comprehension. Most studies investigating multiple document processing use essays as a measure of comprehension (Braasch et al., 2018). Indeed, many writing tasks require reading multiple documents and then writing an essay, report, or response. My current quest is to understand how students comprehend multiple documents, or sources, and how they write about them.

One core endeavor inherent to these objectives is the development of AWEs for source-based writing. We have explored some aspects of multidocument processing using the CNA model. However, we also need to build new algorithms to evaluate constructed responses and provide formative feedback across multiple writing genres, including persuasive essays, essays that rely on single documents (e.g., summaries, recall), and multidocument essays (e.g., source-based essays). In collaboration with Scott Crossley, Laura Allen, and Rod Roscoe, our current objective is to build the Writing Assessment Tool (WAT; IES, R305A180261) to provide an online platform that provides students, teachers, and researchers access to automated writing analytics on persuasive (independent) essays, summaries, and source-based (integrative) essays. WAT will comprise three access points, each tailored to the needs of three types of end-users. *Students* will receive summative and formative feedback via AWE. *Teachers* will have access to a teacher interface allowing them to administer essay assignments, which they can choose to be scored using AWE, grade themselves using scaffolded rubrics, or a combination of the two. *Researchers* will have access to a web-based tool, a downloadable tool, and editable software, which will allow them to conduct computational analyses of writing. In essence, we are building a Coh-Metrix but for writing analytics and for multiple stakeholders. Our overall aim is to provide a writing analytics tool that will enhance students' ability to produce high-quality texts across multiple genres.

Theory. Multiple needs and theories motivate and drive WAT. From the students' perspective, there is a need to increase availability to practice with feedback that enhances students' ability to successfully compose across multiple types of tasks. This returns us to the notion of flexibility. Students need writing practice to learn to recognize and adapt text to varying audiences based on their knowledge, skills, and beliefs and across multiple tasks (e.g., persuasive essays, summaries, source-based essays). Such varied practice is expected to lead to greater flexibility in their use of strategies and approaches to writing as they practice composing, while receiving feedback across a wide variety of contexts (Allen et al., 2016d). Current AWE systems, however, address a very limited number of writing contexts and fail to offer practice opportunities for the wide range of writing tasks that students will likely encounter in academic and professional settings. In turn, most teachers do not have the time or capacity to provide practice with feedback on multiple types of writing and across a sufficient number of writing assignments. WAT cannot replace teachers; such AWE systems are meant to support them to provide automated evaluation, suggest feedback, and understand the features of their students' writing.

One focus of WAT is on source-based essay writing based on multiple documents, which requires the skills to comprehend the documents as well as skills germane to writing. Much of the research and literature that have tackled issues surrounding multidocument comprehension has focused on sourcing, a complex set of competencies that include attending to, representing, evaluating, and using features of information sources such as the author and venue of publication (Braasch et al., 2018; Braasch & Kessler, 2021). Likewise, theoretical accounts of multidocument comprehension emphasize representing source information as a critical component of theoretical frameworks (Rouet et al., 2017) and educational interventions (Braasch et al., 2018). My objective has been to enhance our understanding of comprehension processes when faced with multiple documents. Beyond sourcing, how can we enhance students' comprehension of the relations between documents? The focus of one of my current projects in collaboration with Joe Magliano, Laura Allen, and Kathryn McCarthy is on examining processes engaged during multidocument processing and various ways of improving comprehension, including self-explanation and source evaluation (IES, R305A180144). A related project has focused on integrating reading and writing within iSTART, adding modules that include instruction and practice to better comprehend multiple documents and write source-based essays (ONR, N00014-17-1-2300). Further, in collaboration with Laura Allen, we are investigating how multiple sources of information are processed in social media and in particular how to overcome misinformation and misconceptions (ONR, N00014-19-1-2424).

Our objective across these multiple projects is to better understand which strategies and interventions are most effective in enhancing comprehension of multiple documents and writing essays that reflect integration of the ideas within and across documents. From the perspective of researchers, one of the most common requests regarding access to Coh-Metrix comes from researchers who are seeking to examine features of natural discourse and writing. Coh-Metrix was built to assess text difficulty, not writing quality, and thus it falls short in providing researchers with the necessary linguistic features to capture writing quality (McNamara et al., 2013, 2015). Hence, our objective in the WAT project (IES, R305A180261) is to provide researchers with access to a web-based tool, a downloadable tool, and editable software that will allow researchers to conduct computational analyses of writing. The system will be packaged and disseminated such that researchers and software developers can easily integrate components of WAT into existing tools to provide natural language processing (NLP) extensions in educational systems. Our overall aim is to provide a writing analytics tool that will enhance students' ability to produce high-quality texts across multiple genres. Thus, we are developing a tool that will have a broad impact on current practices in writing research *and* instruction across multiple dimensions.

A multilayered, multidimensional framework

I began my journey with what I believed was a relatively complex notion of comprehension: a connectionist model nested within the complexities of reader and text differences. In the end, with the experiences of developing technologies that target multiple tasks, contexts, and individuals, my own mental model of comprehension has become much more complex. I now realize, of course, that language comprehension and production are intertwined, albeit driven by different skills, contexts, and goals. I have likened aspects of texts, reading, and writing to a diamond because my experiences in research and technology development have led me to conjecture that language production and comprehension is both multilayered and multidimensional (i.e., an M&M Framework[2]).

What do I mean? Of course, it is not challenging to picture layers, simply one layer (of something) on top of another. In this case, each layer might be illustrated using something akin to a connectionist representation. But then, what is the difference between a layer and a dimension?

A layer represents a construct. We evoke constructs. For example, in this article I have evoked constructs relevant to text difficulty, comprehension skills, writing skills, and so on. Even an object such as a *table* is a construct—a table is the idea of an object on which other objects can be placed, and it can be used for various activities such as writing and eating. We naturally evoke the image of a piece of furniture with four legs, but a rock or box can also serve as tables. As such, the notion of a table is a construct because we define it dynamically based on our needs.

In turn, one definition of a dimension is something about an object or shape that can be measured. An example of dimensions in a physical object such as a table are its length, width, and depth. A table has multiple dimensions physically; it also has dimensions related to its functions: what it is for and what we do with it. It is naturally defined using multiple dimensions.

While language is a sequence of words (e.g., table) connected by the glue of syntax, neither words nor syntax are unidimensional. Even a word has multiple dimensions. Of course, the concept that a word has multiple dimensions but the word also has multiple dimensions. For example, the word *table* is a noun, it is a common word, it has many associations and meanings, and so on. Its noun-ness constrains which words and types of words it tends to accompany; its noun-ness constrains how we process it and how we use it differently from other words.

Within the realm of language, we have learned that one layer of language, semantics, has hundreds of dimensions (McNamara, 2011). Latent Semantic Analysis for example, defines semantic spaces using hundreds of dimensions representing different aspects of the relations between words and ideas. Likewise, we have seen that the construct of text difficulty is multidimensional; it varies in terms of multiple components, including narrativity, cohesion, concreteness, and syntax.

At this point, you might imagine that each layer has dimensions (potentially hundreds). Given the construct of layers, we naturally layer one on top of another, like a layer cake. Were it only that simple! By contrast, I'd like to stretch your imagination further by invoking the notion that the layers can cross through other layers, and even share multiple dimensions. Each layer is not separate—they crisscross —sharing elements to give the appearance of a whole. Measuring aspects of language might be viewed as akin to using fractional distillation to separate the components of a liquid. For example, because the liquids in liquid ethanol have different boiling points, it can be separated into ethanol and water. In the realm of language, many researchers have sought to separate out the components: as if a construct such as working memory could be separated from reading skill or semantics could be separated from syntax. However, language is not a chemical solution that can be separated out into its components. Once we separate out one component, we realize that we have missed another one because the two shared common elements comprising their dimensions.

Layers are the constructs that we evoke within theoretical accounts of various phenomena. Our challenge is to find, define, and measure the *layers* of language while at the same time accepting the notion that what we are seeking in language involves multiple constructs (and thus multiple layers) that are sometimes latent and often multidimensional. Language production and comprehension is multi-layered because we tie together information based on multiple aspects of language. Language is multi-dimensional because each layer, or construct, itself cannot be represented entirely with one dimension.

I cannot offer a mathematical solution to this problem beyond the collective work we have conducted in modeling language. What I refer to here as an M&M Framework describes the collective notions of language that have emerged across my research on comprehension and writing and the development of the technologies I and my teams have built to simulate processes related to comprehension and writing. Of course, this entire article might have been organized around a defense of this framework, but that's not what this article was about. It was about the journey, and it was a brief description of where I am now and how the building of technologies has informed how I conceive of language.

Conclusion

A fundamental assumption across all of my projects is that a single researcher cannot have a sufficient and long-lasting impact on a phenomenon and that by providing research tools to the community as a whole, I (personally) have greater potential to have a more profound and lasting impact on our understanding of language and literacy. Research on text and discourse is challenging to conduct due to many factors; one of those factors relates to the time-consuming, arduous, and ultimately high cost of analyzing the language within text, constructed responses, and writing. My hope is that providing researchers with access to technologies will continue to increase and expand research on reading and writing. The tools that we have created and continue to refine facilitate researchers' use of automated text analytics and natural language processing and in turn their capacity to develop automated tutoring technologies.

A second fundamental assumption that currently drives a good deal of my research is that NLP is not just a tool to create algorithms (McNamara et al., 2017, 2018). NLP provides a fundamental means to understand language, comprehension, and communication. Different features of language (e.g., syntax, concreteness, meaningfulness, action, cohesion) provide proxies aligned with how individuals are processing, can process, are producing, and can produce language. To this end, each word, sentence, paragraph, and text is represented with multiple arrays comprising features. For example, the word "chase" can be represented as a verb, its frequency of occurrence in language, its relation to action, associations with other words, and so on. Beyond semantic associations with other words and contexts (cf. Landauer et al., 2007), those features collectively provide multidimensional proxies for meaning (McNamara, 2011).

Accordingly, features of language provide information about communication as well as individual differences in producing and understanding language. For example, if an individual can produce language (e.g., in an essay) with rare words and complex syntax, we can in turn model that individual's skills, such as working memory capacity and reading skill (Allen et al., 2016c). If an individual can produce language that is cohesive and lexically sophisticated, we can predict that the individual is a better reader (Allen et al., 2016a; see also Allen & McNamara, 2015). This is not just a modeling game; this research reflects the underlying assumption that language represents our skills, knowledge, or motivation because features of language reflect how we are processing and producing information.

Within the world of text and discourse, we can quibble about the extent to which reading, writing, and discourse depend on various functions, skills, and processes. Some theories emphasize constructs such as lower level (e.g., lexical) processing and some emphasize higher level (e.g., inferential) processing. It may be helpful, however, to consider those differences within the broader context of other domains. For example, with respect to cognitive psychology, comprehension is generally considered to be a *higher-level process*. By contrast, social psychologists, for example, are likely to consider any cognitive process to be in the realm of *lower-level processes*. Therefore, when we speak of lower-level versus higher-level processing, we often operate within a small range of the full picture. Hence, it is important to not get lost in one's own layer of constructs and to be constrained by our methods and measures. When we put the spotlight on one layer of this complex interplay of mechanisms and processes, it is important to recognize that there are other mechanisms and processes at play as multiple, multidimensional layers of constructs and processes drive language production and comprehension.

Notes

1. Our beloved Carol Connor passed away during the first year of iSTART-Early funding in 2020. Her contributions to the development of the iSTART-Early project were enormous. She is missed.
2. I refer to this as a *framework* rather than a *model* because, as Herb Clark explained to me during his Distinguished Scientific Contribution Award address, a Framework is not intended to be testable; it is a theoretical approach to examining your targeted phenomena.

Acknowledgments

I am extremely grateful to the Society for Text and Discourse, for the honor of granting me the Distinguished Scientific Contributions Award and the mentorship and support from the leaders and members throughout my career, and for the sincere congeniality of the society. I am also sincerely grateful to all of the mentors, colleagues, mentees, and students with whom I have worked with over the past 30+ years; their names appear throughout this tribute. I also extend my gratitude to my friend and colleague, Pani Kendeou, who provided feedback on the first version of this manuscript; Paul Van Den Broek, Morton Ann Gernsbacher, and an anonymous reviewer, who reviewed this article and provided extremely helpful feedback; and the members of my lab, including Tracy Arner, Tong Li, and Natalie Newton, who helped me to put the final touches on this manuscript, in particular the figures, illustrations, and the hopeful absence of typos. I am infinitely grateful to my husband, Grant Sinclair, who has supported me throughout my career in so many ways. Finally, I am grateful to the James S. McDonnell Foundation for jump-starting my career and to my current funding agencies: this article was partially supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305A190063, R305A190050, R305A180144, R305A180261, R305C160004,

R305A150176, and by the Office of Naval Research grants N00014-19-1-2424, N00014-20-1-2623, and N00014-17-1-2300. The opinions expressed are those of the author and do not represent views of the IES, the U.S. Department of Education, or ONR.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Institute of Education Sciences (grants R305A190063, R305A190050, R305A180144, R305A180261, R305C160004, and R305A150176) and Office of Naval Research (grants N00014-20-1-2623, N00014-19-1-2424, and N00014-17-1-2300). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education or the Office of Naval Research.

ORCID

Danielle S. McNamara http://orcid.org/0000-0001-5869-1420

References

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning and Individual Differences*, *6*(4), 379–397. https://doi.org/10.1016/1041-6080(94)90001-9

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014a). Game-based writing strategy tutoring for second language learners: Game enjoyment as a key to engagement. *Language Learning and Technology*, *18*(2), 124–150.

Allen, L. K., Dascalu, M., McNamara, D. S., Crossley, S. A., & Trausan-Matu, S. (2016a). Modeling individual differences among writers using ReaderBench. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *Proceedings of the 8th annual International Conference on Education and New Learning Technologies (EduLearn)* (pp. 5269–5279). Barcelona, Spain: IATED.

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016b). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (*2nd* ed., pp. 316–329). The Guilford Press.

Allen, L. K., Likens, A. D., & McNamara, D. S. (2019). Writing flexibility in argumentative essays: A multidimensional analysis. *Reading and Writing*, *32*(6), 1607–1634. https://doi.org/10.1007/s11145-018-9921-y

Allen, L. K., & McNamara, D. S. (2015). You are your words: Modeling students' vocabulary knowledge with natural language processing. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th international conference on educational data mining (EDM 2015)* (pp. 258–265). Madrid, Spain: International Educational Data Mining Society.

Allen, L. K., Perret, C. A., & McNamara, D. S. (2016c). Linguistic signatures of cognitive processes during writing. In J. Trueswell, A. Papafragou, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 38th annual meeting of the cognitive science society in Philadelphia, PA* (pp. 2483–2488). Austin, TX: Cognitive Science Society.

Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014b). Reading comprehension components and their relation to the writing process. *L'année psychologique/Topics in Cognitive Psychology*, *114*(4), 663–691. https://doi.org/10.4074/S0003503314004047

Allen, L. K., Snow, E. L., & McNamara, D. S. (2016d). The narrative waltz: The role of flexibility on writing performance. *Journal of Educational Psychology*, *108*(7), 911–924. https://doi.org/10.1037/edu0000109

Anderson, J. R. (1990). *The adaptive character of thought*. Erlbaum.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (M. Holquist, Ed., C. Emerson, & M. Holquist, Trans.). University of Texas Press.

Balyan, R., McCarthy, K. S., & McNamara, D. S. (2020). Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education (IJAIED)*, *30*(3), 337–370. https://doi.org/10.1007/s40593-020-00201-7

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, *25*(1), 65–83. https://doi.org/10.1097/00011363-200501000-00007

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, *13*(2), 221–252. https://doi.org/10.1207/s1532690xci1302_3

Boonthum, C., Levinstein, I., & McNamara, D. S. (2007). Evaluating self-explanations in iSTART: Word matching, latent semantic analysis, and topic models. In A. Kao & S. Poteet (Eds.), *Natural language processing and text mining* (pp. 91–106). Springer-Verlag UK.

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

Braasch, J. L. G., & Kessler, E. D. (2021). Towards a theoretical model of source comprehension in everyday discourse. *Discourse Processes*, 1–19. (note this paper is in the same special issue). https://doi.org/10.1080/0163853X.2021.1905393

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

Cioaca, V., Dascalu, M., & McNamara, D. S. (2020). Extractive summarization using cohesion network analysis and submodular set functions. In *22nd international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)* (pp. 161–168). Timisoara, Romania: IEEE.

Crossley, S. A., Allen, L. K., & McNamara, D. S. (2016a). The Writing Pal: A writing strategy tutor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 204–224). Routledge.

Crossley, S. A., Kim, M., Allen, L. K., & McNamara, D. S. (2019). Automated summarization evaluation (ASE) using natural language processing tools. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Proceedings of the 20th international conference of artificial intelligence in education (AIED), lecture notes in computer science* (vol. 11625, pp. 84–95). Chicago, IL: Springer.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *The Journal of Writing Assessment*, *8*(1), 1–16. https://www.researchgate.net/publication/319604495_To_Aggregate_or_Not_Linguistic_Features_in_Automatic_Essay_Scoring_and_Feedback_Systems

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., & McNamara, D. S. (2014). Developing component scores from natural language processing tools to assess human ratings of essay quality. In W. Eberle & C. Boonthum-Denecke (Eds.), *Proceedings of the 27th international Florida artificial intelligence research society (FLAIRS) conference* (pp. 381–386). Palo Alto, CA: AAAI Press.

Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write high quality essays. *Written Communication*, *31*(2), 181–214. https://doi.org/10.1177/0741088314526354

Crossley, S. A., Sirbu, M. D., Dascalu, M., Barnes, T., Lynch, C. F., & McNamara, D. S. (2018). Modeling math success using cohesion network analysis. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. D. Boulay (Eds.), *Proceedings of the 19th international conference on artificial intelligence in education (AIED 2018), part II* (pp. 63–67). London, UK: Springer.

Dascalu, M., Crossley, S. A., McNamara, D. S., Dessus, P., & Trausan-Matu, S. (2018a). Please Readerbench this text: A multi-dimensional textual complexity assessment framework. In S. Craig (Ed.), *Tutoring and intelligent tutoring systems* (pp. 251–271). Nova Science Publishers, Inc. https://research.ou.nl/en/publications/please-readerbench-this-text-a-multi-dimensional-textual-complexi

Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L. K. (2018b). Cohesion network analysis of CSCL participation. *Behavior Research Methods*, *50*(2), 604–619. https://doi.org/10.3758/s13428-017-0888-4

Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench – Automated evaluation of collaboration based on cohesion and dialogism. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 395–423. https://doi.org/10.1007/s11412-015-9226-y

Dascalu, M. D., Ruseti, S., Carabas, M., Dascalu, M., Trausan-Matu, S., & McNamara, D. S. (2020a). Cohesion network analysis: Predicting course grades and generating sociograms for a Romanian moodle course. In *16th international conference on intelligent tutoring systems (ITS 2020)*. Athens, Greece: Springer.

Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., & Trausan-Matu, S. (2020b). Multi-document cohesion network analysis: Visualizing intratextual and intertextual links. In I. Ibert Bittencourt, M. Cukorova, K. Muldner, E. Millan, & R. Luckin (Eds.), *Proceedings of the 21st international conference on artificial intelligence in education (AIED 2020)*. Ifrane, Morocco: Springer.

Doane, S. M., McNamara, D. S., Kintsch, W., Polson, P. G., & Clawson, D. M. (1992). Prompt comprehension in UNIX command production. *Memory & Cognition*, *20*(4), 327–343. https://doi.org/10.3758/BF03210918

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363. https://doi.org/10.1037/0033-295X.100.3.363

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*(2), 371–398. https://doi.org/10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R., & the Tutoring Research Group. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, *1*(1), 35–51. https://doi.org/10.1016/S1389-0417(99)00005-4

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist*, *53*(4), 258–279. https://doi.org/10.1080/00461520.2018.1481406

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 383–402). Guilford Press.

Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, *81*(4), 710–744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., Crutcher, R. J., King, C., Ericsson, K. A., & Bourne, L. E., Jr. (1993). The long-term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation* (pp. 135–164). Academic Press.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual Review of Psychology*, *47*(1), 143–172. https://doi.org/10.1146/annurev.psych.47.1.143

Healy, A. F., Oliver, W. L., & McNamara, T. P. (1987). Detecting letters in continuous text: Effects of display size. *Journal of Experimental Psychology Human Perception and Performance*, *13*(2), 279–290. https://doi.org/10.1037/0096-1523.13.2.279

Healy, A. F., Schneider, V. I., & Bourne, L. E., Jr. (2012). Empirically valid principles of training. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Training cognition: Optimizing efficiency, durability, and generalizability* (pp. 13–39). Psychology Press.

Jackson, T. G., Boonthum, C., & McNamara, D. S. (2015). Natural language processing and game-based practice in iSTART. *Journal of Interactive Learning Research*, *26*(2), 189–208. https://www.learntechlib.org/p/42009/

Jackson, T. G., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 1036–1049. https://doi.org/10.1037/a0032580

Jackson, T. G., & McNamara, D. S. (2017). The motivation and mastery cycle framework: Predicting long-term benefits of educational games. In Y. Baek (Ed.), *Game-based learning: Theory, strategies and performance outcomes* (pp. 97–122). Nova Science Publishers.

Johnson, A. M., Guerrero, T. A., Tighe, E. L., & McNamara, D. S. (2017). iSTART-ALL: Confronting adult low literacy with intelligent tutoring for reading comprehension. In B. Boulay, R. Baker, & E. Andre (Eds.), *Proceedings of the 18th international conference on artificial intelligence in education (AIED)* (pp. 125–136). Wuhan, China: Springer.

Johnson, A. M., Perret, C. A., Watanabe, M., Kopp, K., McCarthy, K. S., & McNamara, D. S. (2018). Adaptive literacy instruction in iSTART and W-Pal: Implementing the outer loop. In S. Craig (Ed.), *Tutoring and intelligent tutoring systems* (pp. 221–250). Nova Science Publishers.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective and Behavioral Neuroscience*, *7*(3), 153–191. https://doi.org/10.3758/CABN.7.3.153

Kintsch, W. (1988). The role of knowledge in discourse comprehension construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

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

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Erlbaum.

Levinstein, I. B., Boonthum, C., Pillarisetti, S. P., Bell, C., & McNamara, D. S. (2007). iSTART 2: Improvements for efficiency and effectiveness. *Behavior Research Methods*, *39*(2), 224–232. https://doi.org/10.3758/BF03193151

MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.). (2006). *Handbook of writing research*. Guilford.

Magliano, J. P., Todaro, S., Millis, K. K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, *32*(2), 185–208. https://doi.org/10.2190/1LN8-7BQE-8TN0-M91L

McCarthy, K. S., Likens, A. D., Johnson, A. M., Guerrero, T. A., & McNamara, D. S. (2018). Metacognitive overload! Positive and negative effects of metacognitive prompts in an intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, *28*(3), 420–438. https://doi.org/10.1007/s40593-018-0164-5

McCarthy, K. S., Watanabe, M., Dai, J., & McNamara, D. S. (2020b). Personalized learning in iSTART: Past modifications and future design. *Journal of Research on Technology in Education*, *52*(3), 301–321. https://doi.org/10.1080/15391523.2020.1716201

McCrudden, M. T., & McNamara, D. S. (2017). *Cognition in education*. Routledge.

McNamara, D. S. (1995). Effects of prior knowledge on the generation advantage: Calculators versus calculation to learn simple multiplication. *Journal of Educational Psychology*, *87*(2), 307–318. https://doi.org/10.1037/0022-0663.87.2.307

McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*(1), 51–62. https://doi.org/10.1037/h0087352

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S. (2009). The importance of teaching reading strategies. *Perspectives on Language and Literacy*, *35*, 34–40. https://www.dropbox.com/s/60hmkbyq82n36zu/152_Teaching%20Reading%20Strategies%20-%20McNamara.pdf?dl=0

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, *3*(1), 3–17. https://doi.org/10.1111/j.1756-8765.2010.01117.x

McNamara, D. S. (2013). The epistemic stance between the author and reader: A driving force in the cohesion of text and writing. *Discourse Studies*, *15*(5), 575–592. https://doi.org/10.1177/1461445613501446

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*, *54*(7), 479–492. https://doi.org/10.1080/0163853X.2015.1101328

McNamara, D. S. (2020). If integration is the keystone of comprehension: Inferencing is the key. *Discourse Processes*, *58*(1), 86–91. https://doi.org/10.1080/0163853X.2020.1788323

McNamara, D. S., & Allen, L. K. (2017). Toward an integrated perspective of writing as a discourse process. In M. Schober, A. Britt, & D. N. Rapp (Eds.), *Handbook of discourse processes* (*2nd* ed., pp. 362–389). Routledge.

McNamara, D. S., Allen, L. K., Crossley, S. A., Dascalu, M., & Perret, C. A. (2017). Natural language processing and learning analytics. In G. Siemens & C. Lang (Eds.), *Handbook of learning analytics and educational data mining* (pp. 93–104). Society for Learning Analytics Research.

McNamara, D. S., Allen, L. K., McCarthy, K. S., & Balyan, R. (2018). NLP: Getting computers to understand discourse. In K. Millis, D. Long, J. Magliano, & K. Wiemer (Eds.), *Deep comprehension: Multi-disciplinary approaches* (pp. 224–236). Routledge.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007a). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 227–241). Erlbaum.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010a). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*(2), 499–515. https://doi.org/10.3758/s13428-012-0258-1

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35–59. https://doi.org/10.1016/j.asw.2014.09.002

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89–116). Rowman & Littlefield Education.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

McNamara, D. S., & Healy, A. F. (1995a). A generation advantage for multiplication skill and nonword vocabulary acquisition. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills* (pp. 132–169). Sage.

McNamara, D. S., & Healy, A. F. (1995b). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language*, *34*(3), 399–416. https://doi.org/10.1006/jmla.1995.1018

McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2010b). 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). IGI Global.

McNamara, D. S., Jacovina, M. E., Snow, E. L., & Allen, L. K. (2015b). From generating in the lab to tutoring systems in classrooms. *American Journal of Psychology*, *128*(2), 159–172. https://doi.org/10.5406/amerjpsyc.128.2.0159

McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, *4*(1), 33–46. https://www.researchgate.net/publication/281894454_Translating_advances_in_reading_comprehension_research_to_educational_practice

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–287. https://doi.org/10.1080/01638539609544975

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, and Computers*, *36*(2), 222–233. https://doi.org/10.3758/BF03195567

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297–384). Elsevier.

McNamara, D. S., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In A. M. Columbus (Ed.), *Advances in psychology research* (pp. 1–24). Nova Science Publishers, Inc.

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with Istart. *Journal of Educational Computing Research*, *34*(2), 147–171. https://doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE

McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007b). 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). Erlbaum.

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, *4*(1), 229–257. https://www.pegem.net/dosyalar/dokuman/138547-2014010711325-14.pdf

Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., & McNamara, D. S. (2007). Assessing and improving comprehension with latent semantic analysis. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 207–225). Erlbaum.

Nicula, B., Perret, C. A., Dascalu, M., & McNamara, D. S. (2019). Predicting multi-document comprehension: Cohesion network analysis. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Proceedings of the 20th international conference of artificial intelligence in education (AIED)* (pp. 358–369). Chicago, IL: Springer.

Nicula, B., Perret, C. A., Dascalu, M., & McNamara, D. S. (2020). Extended multi-document cohesion network analysis centered on comprehension prediction. In I. Ibert Bittencourt, M. Cukorova, K. Muldner, E. Millan, & R. Luckin (Eds.), *Proceedings of the 21st international conference on artificial intelligence in education (AIED 2020)*. Ifrane, Morocco: Springer.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*(2), 121–152. https://doi.org/10.1080/01638530709336895

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Paraschiv, I. C., Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Banica, C. K. (2017). Exploring the LAK dataset using cohesion network analysis. In D. Trandabat & D. Gifu (Eds.), *3rd workshop on social media and the web of linked data (RUMOUR 2017), in conjunction with the joint conference on digital libraries (JCLD 2017)* (pp. 17–21). "Alexandru Ioan Cuza" University Publishing House.

Proske, A., Roscoe, R. D., & McNamara, D. S. (2014). Game-based practice versus traditional practice in computer-based writing strategy training: Effects on motivation and achievement. *Education Technology Research Development*, *62*(5), 481–505. https://doi.org/10.1007/s11423-014-9349-2

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). Academic Press.

Roscoe, R. D., Allen, L. K., & McNamara, D. S. (2019). Contrasting writing practice formats in a writing strategy tutoring system. *Journal of Educational Computing Research*, *57*(3), 723–754. https://doi.org/10.1177/0735633118763429

Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, *34*, 39–59. https://doi.org/10.1016/j.compcom.2014.09.002

Roscoe, R. D., Jacovina, M. E., Harry, D., Russell, D. G., & McNamara, D. S. (2015). Partial verbal redundancy in multimedia presentations for writing strategy instruction. *Applied Cognitive Psychology*, *29*(5), 669–679. https://doi.org/10.1002/acp.3149

Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, *105*(4), 1010–1025. https://doi.org/10.1037/a0032340

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015b). Automated detection of essay revising patterns: Application for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition, and Learning*, *10*(1), 59–79. https://files.eric.ed.gov/fulltext/ED565460.pdf

Roscoe, R. D., Snow, E. L., Brandon, R. D., & McNamara, D. S. (2013a). Educational game enjoyment, perceptions, and features in an intelligent writing tutor. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th international Florida artificial intelligence research society (FLAIRS) conference* (pp. 515–520). Menlo Park, CA: AAAI Press.

Roscoe, R. D., Varner (Allen), L. K., Crossley, S. A., & McNamara, D. S. (2013b). Developing pedagogically-guided threshold algorithms for intelligent automated essay feedback. In *International journal of learning technology* (Vol. 8, pp. 362–381). Int. J. Learning Technology. https://files.eric.ed.gov/fulltext/ED585772.pdf

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*(2), 181–221. https://doi.org/10.3102/00346543066002181

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015

Ruseti, S., Dascalu, M., Johnson, A., McNamara, D. S., Balyan, R., Kopp, K., Crossley, S. A., & Trausan-Matu, S. (2018b). Predicting question quality using recurrent neural networks. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. D. Boulay (Eds.), *Proceedings of the 19th international conference on artificial intelligence in education (AIED 2018), part I* (pp. 491–502). London, UK: Springer.

Ruseti, S., Dascalu, M., Johnson, A. M., McNamara, D. S., Balyan, R., McCarthy, K. S., & Trausan-Matu, S. (2018a). Scoring summaries using recurrent neural networks. In R. Nkambou, R. Azevedo, & J. Vassileva (Eds.), *Proceedings of the 14th international conference on intelligent tutoring systems (ITS)* (pp. 191–201). Montreal, Canada: Springer.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97–118). Cambridge University Press.

Snow, E. L., Jacovina, M. E., Jackson, G. T., & McNamara, D. S. (2016). iSTART-2: A reading comprehension and strategy instruction tutor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 104–121). Routledge.

Watanabe, M., McCarthy, K., & McNamara, D. S. (2019). Examining the effects of adaptive task selection on students' motivation in an intelligent tutoring system. In S. Hsiao, J. Cunningham, K. McCarthy, G. Lynch, N. Hoover, C. Brooks, R. Ferguson, & U. Hoppe (Eds.), *Companion proceedings of the 9th international conference on learning analytics and knowledge (LAK'19)* (pp. 161–162). Phoenix, AZ: SOLAR.