Intelligent Tutoring Systems for Literacy: Existing Technologies and Continuing Challenges

Matthew E. Jacovina and Danielle S. McNamara

Arizona State University

Author's Note

This work was supported by the Institute of Education Sciences (IES), US Department of Education, through grants R305A120707 and R305A130124, and by the Office of Naval Research (ONR) through grant N000141410343. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily represent views of the IES or the ONR. We also thank Laura Allen for her helpful comments.
Completed 2017.

Jacovina, E. J., & McNamara, D. S. (in press). Intelligent tutoring systems for literacy: Existing technologies and continuing challenges. In R. Atkinson (Ed.), *Intelligent tutoring systems: Structure, applications and challenges*. Hauppauge, NY: Nova Science Publishers Inc. To be published with acknowledgment of federal support.

Abstract

In this chapter, we describe several intelligent tutoring systems (ITSs) designed to support student literacy through reading comprehension and writing instruction and practice. Although adaptive instruction can be a powerful tool in the literacy domain, developing these technologies poses significant challenges. For example, evaluating the quality of a student's writing can be challenging because of the numerous ways to succeed (or fail) when generating a written work. Throughout our discussion, we focus on the methodologies that ITSs have employed to face these challenges. Natural language processing techniques, for example, can be leveraged to assess students' level of comprehension or writing proficiency and subsequently drive the feedback that students receive. Additional challenges arise in the implementation of these systems in classrooms; we discuss how the features and flexibility offered by ITSs can augment their usefulness in these real-world settings. We conclude the chapter by forecasting how future generations of ITSs for literacy will improve and fit into the educational landscape.

Intelligent Tutoring Systems for Literacy: Existing Technologies and Continuing Challenges

Literacy skills are vital for success in modern occupations and are becoming increasingly important for social communication. For decades, researchers and developers have worked to create and improve educational technologies to provide and supplement literacy instruction (Cheung & Slavin, 2012). These technologies run the spectrum from teaching decoding skills to promoting higher-level argumentation in academic writing. Providing student-specific feedback can enhance the effectiveness of tutoring systems by focusing students on pertinent materials and providing appropriate feedback and scaffolding (e.g., Kegel & Bus, 2012). Hence, making these technologies *adaptive* to individual students and their needs is at the forefront of the goals faced by intelligent tutoring systems (ITSs).

An overarching challenge for these systems is inherent to the assessment of literacy skills: there are often multiple solutions for a given problem (e.g., when responding to an essay prompt) and there is a level of subjectivity to that assessment. Reading and writing are sometimes referred to as *ill-defined* domains in the ITS literature – a term that is suggestive of the multiple challenges in building reading and writing ITSs (Fournier-Viger, Nkambou, & Nguifo, 2010; Le, Loll, & Pinkwart, 2013; Lynch, Ashley, Pinkwart, & Aleven, 2009). Whereas an item in a *well-defined* domain such as math may have only one correct answer (e.g., 18/2=9), even a simple writing prompt may have numerous approaches and content that could lead to success. Hence, compared to well-defined domains, ill-defined domains raise different challenges when building an expert model and tracking students' knowledge states through a student model.

In this chapter, we describe several technologies for literacy instruction. Throughout, we highlight the challenges these technologies face and how they attempt to overcome or circumvent them. Not all of the systems described in this chapter meet traditional definitions of an ITS (VanLehn, 2006). Although there is value in pursuing the full suite of advantages of an ITS within literacy domains (Neuwirth, 2014), many positive outcomes have been obtained with computer-based tutors that employ only modest amounts of student modeling and levels of adaptivity. ITSs in some domains have used techniques such as Bayesian Knowledge-Tracing to track students' mastery of skills, or knowledge components, over time (e.g., Desmarais & Baker, 2012). While successful in well-defined domains such as geometry or algebra, it is difficult and, may be impossible, to trace individual knowledge components related to literacy while students are learning to read and write. Reading and writing are complex skills that require the integration of multiple declarative and procedural components to complete tasks. Moreover, the successful execution of these skills is highly contingent on the literary context. For example, one crucial skill in reading involves generating bridging inferences to connect ideas across different parts of text or discourse. A reader might be able to generate bridging inferences in a simple text describing the steps in making a sandwich, but not while reading a text that describes human digestion. Likewise, while writing, a student may successfully compose the introduction of an essay that discusses the impact of the arts (literature, music, painting) on society, but falter when faced with a prompt on the impact of culture on changes in the arts across history.

Various techniques have been used by ITS developers to adapt instruction to students in the absence of discrete measures of particular skills. One approach to automating one-on-one feedback and adapting instruction comes from the use of natural language processing (NLP). NLP is the analysis of naturally occurring human language by a computer. Several systems reviewed in this chapter use NLP techniques to automatically analyze the language used in students' responses and provide meaningful, adaptive feedback (Boonthum-Denecke, Levinstein, McNamara, Magliano, & Millis, 2008; McNamara, Crossley, & Roscoe, 2013). The goal for NLP in the context of automated tutoring of reading or writing is to replicate some qualities of teacher-given feedback, be it summative assessments of quality, or formative feedback intended to improve the future quality of students' work.

When teachers provide such feedback to students' written responses, they assess the quality of students' discourse (e.g., an essay or an answer to a comprehension question), which is a complex and time consuming process, and provide summaries of this assessment and sometimes suggestions for improvement. This feedback can encourage high-quality, deliberate practice from students, which is important for mastering complex skills (Ericsson, 2008; Ericsson, Krampe, & Tesch-Römer, 1993). Perfectly automating this teacher-given feedback is currently not possible, at least not across a sizeable range of contexts. However, NLP can be used to build models linking linguistic features of students' responses to measures of quality and the need for particular kinds of feedback messages. Subsequently, these models can automatically assess students' written responses. As such, the linguistic features in the essays provide *proxies* to the knowledge components assumed to comprise the skills essential to successful writing (McNamara, Graesser, McCarthy, & Cai, 2013).

The selection of technologies described in this chapter is not meant to be exhaustive, nor do our descriptions cover the breadth of the features and research behind each system. Each of the systems in this chapter (along with several others) is described more fully in the upcoming book, *Adaptive Educational Technologies for Literacy Instruction* (Crossley & McNamara, in press). Our goal for providing these short summaries is twofold. First, we wish to broadcast that there are a range of tutoring systems available for literacy instruction in the classroom. Although some of these projects have been under development for many years and have already been adopted in several schools, others were developed more recently and have not been widely adopted (yet). When appropriate, we also describe how the systems fit into the classroom and how they provide support to teachers. One of the primary motivations for building literacy technologies is to provide students with individualized instruction and feedback in classrooms as a supplement to teacher-directed instruction. Hence, systems should be developed such that they are easily integrated into classrooms, with adequate supports in place for teachers.

The second goal of this chapter is to summarize the types of adaptation that are currently being used in literacy tutors. Despite the challenges mentioned for adaptive instruction, several clever individualizing features have been implemented successfully. Thus, these summaries may alert other ITS developers to the strengths (and weaknesses) of current and emerging systems. We conclude the chapter by previewing a selection of emerging systems, and by tying together emerging themes, attempt to forecast future directions of educational technologies for literacy.

Reading-Focused Technologies

The following four technologies focus primarily on reading instruction and practice. Reading is a process that involves lower level skills, such as the ability to decode and understand words, and higher level skills, such as the ability to make inferences that bridge information across a text (see McNamara & Magliano, 2009). Readers *comprehend* a text when they are able to extract information from it using a combination of these higher and lower level skills to form a mental representation that combines information from the text with prior knowledge. The systems described in this section focus on different skills that contribute to the comprehension process. These systems range from providing vocabulary training (DSCoVAR), to providing comprehension training (ITSS and iSTART-2), to making assessments of reading skills that lead directly to recommendations to teachers (A2i).

The Dynamic Support of Contextual Vocabulary Acquisition for Reading (DSCoVAR)

The DSCoVAR system is designed to support 4^{th}-8^{th} grade students' vocabulary knowledge through contextual word learning (Frishkoff et al., in press). DSCoVAR focuses on Tier 2 words, which are commonly used across different text domains, but that are not used frequently in speech; understanding many Tier 2 words thus supports comprehension. Moreover, the ability to ascertain the meanings Tier 2 words through context clues is important for students’ ability to understand texts in less familiar domains. Students using DSCoVAR encounter new vocabulary words within a given context (i.e., a sentence or passage) and are asked to type a similar word into a prompt. The corpora include contexts that provide either no cues, limited cues, or strong cues to the meaning of the targeted word. The active learning task encourages students to make inferences about the word meaning when providing their answer. DSCoVAR also provides students with strategy instruction on *how* to use context to arrive at a word definition. This feature can thus be leveraged by students who struggle with the task, making DSCoVAR a more complete package for vocabulary learning.

DSCoVAR is currently able to provide some adaptivity and the developers are working toward incorporating more robust adaptivity. The system provides feedback that extends beyond a simple correct/incorrect distinction. Partial knowledge of a word can be detected by DSCoVAR using NLP techniques (Markov Estimation of Semantic Association; Collins-Thompson & Callan, 2007) to detect the similarity between the student-entered word and the targeted vocabulary word. The authors note that because the system allows students to generate their own answers, students might frequently guess incorrect or only partially correct answers. The ability to provide more nuanced feedback messages is thus a particularly important way to adapt to individual students. The system can also begin students with context cues that make target words relatively easy to figure out, and gradually move to less informative cues, encouraging students to work harder after practicing with basic items. This feature will be improved in future versions such that performance will influence the amount of scaffolding required over time. A bigger challenge will be increasing the size of the corpus, which was built by experts in a time-intensive process. The authors are optimistic, however, that NLP techniques can be leveraged to automatically build a database of additional target words and contexts.

Intelligent Tutoring of the Structure Strategy (ITSS)

ITSS provides reading comprehension instruction through structure strategy training and is intended for 4th to 8th grade students (Meyer & Wijekumar, in press). The structure strategy is designed to help students generate coherent mental representations for expository and argumentative texts. Students engage in several activities, including identifying signaling words, and making text structure classifications (e.g., comparison, problem and solution, cause-and-effect, or description). The idea is that these activities help students impose structure on the texts they read. When students are accustomed to various text structures, they are better able to make connections between ideas in a text—for example, when students encounter a problem, they will know to look for a solution. Making such connections within texts increases students' understanding and ability to recall information. Based on the nature of the particular text structure, students also write main ideas using scaffolds appropriate for that structure (e.g., a comparison structure is scaffolded by prompting students to write what ideas are being compared and on what basis). ITSS also provides video tutorials for teacher professional development, supporting its use in classrooms.

ITSS includes over 100 interactive lessons on the structure strategy. These lessons include researcher-created passages as well as authentic texts, and an animated tutor assists students in reading some of the practice texts. For scoring purposes, each text is broken down into signaling words, main ideas, and details. The responses that students generate during lesson activities are automatically scored using these classifications. For example, when identifying main ideas, a student's response is cleaned (using a spell checker) and compared to the text's list of main ideas, with synonyms counting as hits (Meyer, Wijekumar, & Lin, 2011). The student is then provided feedback based on the percentage of main ideas generated. In addition, game-based activities within ITSS provide a review of the skills and strategies in a less demanding (and potentially more fun) format.

Several features within ITSS have been empirically studied and used to inform decisions about the system. For example, one study found evidence that more elaborate feedback was more helpful to students than simple feedback messages (Meyer et al., 2010). In another study, an adaptive version of ITSS was found to be more successful than the standard ITSS (Meyer et al., 2011). In the adaptive system, performance on the current lesson influenced the selection of the next lesson; for example, poor performance might lead to the next lesson including a text with a lower reading level. The adaptive features of ITSS have allowed for a more personalized user experience, despite the absence of computational complexity. The cost has been the time-consuming nature of developing quality lesson content and practice materials—which, given the successes of ITSS, has been worthwhile.

Interactive Strategy Tutor for Active Reading and Thinking -2 (iSTART-2)

iSTART-2 is a web-based tutoring system that provides reading comprehension instruction and practice for students in middle school through college (Snow, Jacovina, Jackson, & McNamara, in press; Jackson & McNamara, 2013). The system instructs students on strategies to better self-explain difficult texts during reading, including strategies to generate bridging inferences and elaborations (McNamara, 2004). These self-explanations help support deep comprehension of texts, which is especially important for complex technical content, such as science texts. The instructional lessons are presented by a pedagogical agent and present examples of the strategies being used in response to texts. Checkpoint questions that follow lessons assess students' memory for the strategies. After completing the instructional lessons, students are transitioned to a game-based practice interface.

In the practice interface, students can engage with mini-games in which they are presented with self-explanations ostensibly written by other students and asked to identify which of the iSTART-2 strategies were used to create that self-explanation. In generative games, students read a text and type self-explanations in response to target sentences. These self-explanations are automatically scored by an algorithm that uses a combination of word-based approaches and latent semantic analysis to assign a score from 0 to 3 (LSA: Landauer, McNamara, Dennis, & Kintsch, 2007; Jackson, Guess, & McNamara, 2010). Higher scores are assigned to self-explanations that incorporate information from throughout the text and prior knowledge from outside the text, whereas lower scores are assigned to self-explanations that are short, irrelevant, or too similar to the target sentence. Word-based approaches includes matching content words in a student's self-explanation in the text, which can help detect when students, for example, are copying most or all of the target sentence in their self-explanation. LSA provides a way to calculate the semantic overlap between the target sentence and the self-explanation, as well as the previous sections of the text and the student's self-explanation. When there is similarity between the previous text and the self-explanation, the student is likely *bridging* between what came earlier and the current sentence, which corresponds to one of the iSTART-2 strategies. The goal of the algorithm is to assess the quality of the self-explanation in terms of how well a student followed the strategies taught by iSTART-2; the algorithm does not score on the *accuracy* of the content in the self-explanations. This makes the algorithm flexible, allowing teachers to input their own texts into the system. Feedback mechanisms in certain activities provide both a score and suggestions on how to improve when appropriate. For example, when students' scores are low because they are not including information from outside the target sentence, an agent will encourage students to bridge to earlier parts of the text or to elaborate using prior knowledge.

The NLP techniques employed by the iSTART-2 scoring algorithm give students the opportunity to practice self-explaining while receiving automatic feedback. This is an important feature that greatly reduces the amount of one-on-one tutoring required in self-explanation training (McNamara, 2004). The ability for teachers to insert their own texts into the system and have their students receive automated feedback is also a key feature that takes advantage of NLP techniques. Although iSTART-2 comprehension training and practice is useful even when the content does not match exactly with a course's curriculum, teachers of course prefer an alignment. NLP techniques thus afford flexibility and the ease of content creation that make iSTART-2 attractive for classroom implementation.

Assessment-to-Instruction (A2i)

Unlike the other technologies described in this chapter, teachers are the primary users of A2i (Connor et al., 2013; Ingebrand & Connor, in press). The overall goal of the system is to provide teachers with information about what students currently know and what activities would be most appropriate for them to continue advancing, supporting individualized student instruction (ISI). The system has been primarily used with kindergarten through third grade students. Teachers receive recommendations on how much time students should spend per day/week in two types of activities: code-focused or meaning-focused. Code-focused activities center on decoding skills, such as phonological awareness. Meaning-focused activities encourage students to construct knowledge from texts, such as through comprehension strategy instruction. The system also recommends whether these activities should be conducted in teacher-led groups or if a student can practice the skills individually or with a group of peer learners. Teachers can request for A2i to recommend a certain number of teacher-led groups and the system will attempt to cluster students who have similar needs. In this way, A2i helps teachers make decisions about how to manage class time effectively. A2i also provides teachers with professional development that can help them best use the system and follow its recommendations.

To make its recommendations, A2i administers formative assessments to students for word knowledge, decoding, and comprehension. Item response theory (IRT) analyses allow these assessments to be adaptive to students' knowledge level, reducing the administration time. Results from these assessments then form the basis for the recommendations that teachers receive regarding students' optimal practice trajectories. Throughout the school year, students may retake these assessments to update their set of recommendations. The algorithm that drives these recommendations is based on hierarchical linear models (HLM) that predict reading growth based on the month of the school year, amount of practice, and current literacy levels (see Connor et al., 2009 for details). These algorithms have been updated using results from additional studies, thus optimizing the accuracy of the recommendations.

The methods employed in A2i showcase the ability for literacy tutors to provide intelligent recommendations for teachers. A2i's successes suggest the importance of breaking down comprehension skills into components that can be individually targeted for particular students. The backbone for A2i is the abundance of quality research on how readers develop and the pedagogical techniques that can support that development. As theories and models of comprehension in more advanced readers and writers continue to link to educational practices, similar technologies should emerge. Another crucial part of A2i's success is how it integrates teachers into its goals. Although most technologies assume some teacher support, teachers directly provide the instruction when using A2i. However, A2i is not a simple list of activities for a teacher to deliver; it is sensitive to the needs of students and the teacher.

Writing-Focused Technologies

The following four technologies focus primarily on writing instruction and practice. Like reading, writing is a complex process that requires the combination of a number of skills, including the ability to engage in critical thinking, knowledge about conventions of writing, and flexibility to apply these skills in a variety of domains (Framework for Success in Postsecondary Writing, 2011). The systems described focus on formal composition, such as writing persuasive essays or scientific texts. Although other forms of written communication (e.g., Facebook, text messages, etc.) are important, more formal writing is crucial for success in both academic settings and the workforce. The systems that we review in this chapter range from providing automated feedback to students' writing (Criterion and Writing Pal), to supporting peer review (SWoRD), to providing instruction and exemplars for advanced academic writing (RTW).

Criterion

The Education Testing Service's (ETS) Criterion system uses NLP techniques to score and provide feedback to students' writing (Leahy et al., 2014; Ramineni & Deane, in press). A primary goal for Criterion is to provide useful feedback to students on multiple drafts of an essay and to do so without creating an overwhelming burden on teachers to score numerous pieces of writing. The system includes over 400 expository and argumentative prompts ranging from grade levels 4 through college. Digital tools are provided to writers to help them craft their essay, such as the ability to generate outlines and spelling and grammar checks. Feedback can be provided for multiple revisions, and a final draft can be submitted to a teacher for formal review. Several tools are also available to teachers, such as the ability to track students' progress and view essay drafts.

The scoring and feedback within Criterion is driven by the *e-rater* scoring engine (for details, see Deane, 2013; Ramineni & Williamson, 2013). Building a scoring model involves first collecting a large corpus of expert-scored essays, which is then divided into a model building set and an evaluation set. Linguistic features of the essays are extracted from the training set and are regressed onto the essay scores, determining the weights for each feature.

The features included are updated frequently, often depending on the state of the art in NLP. For example, two features that relate to students' vocabulary are average word length and word frequency. The scoring models can be built on a per-prompt basis, or across a number of different prompts to create a generic model. Both methods have advantages and disadvantages: prompt-specific scoring models allow content to be considered in the scoring (e.g., having certain key vocabulary words), but these models are inflexible. Generic models sacrifice some accuracy in order to be more flexible and allow custom prompts to be assigned by the teacher. During the evaluation phase, the resulting model is conducted, with the most obvious step being to score the evaluation set using the scoring model. Higher agreement with the expert-scored essays denotes a successful model. Other criteria are also considered in the evaluation phase, such as whether the scoring model is more accurate for certain subgroups of students. Different scoring models are created for different grade levels and sometimes for specific prompts, and the resulting scores are displayed to students and teachers. Along with a holistic score, students also receive feedback on language errors (e.g., grammatical errors) and discourse elements (e.g., the absence of a thesis statement).

Criterion's approach to providing individualized scoring and feedback to students has clear advantages and upside. As the scoring and feedback algorithms are improved, it could become possible to provide thorough, personalized feedback. Although ETS recommends that Criterion supplement writing instruction rather than replacing teacher feedback, its success suggests that their approach is a good one. A key remaining challenge is that the more flexible "generic" scoring models cannot currently provide the same types of feedback as the prompt-specific models, and even the prompt-specific models cannot accurately evaluate the *quality* of the content within an argument. Although some elements of automated essay scoring are likely to improve in the near future with the advent of new NLP tools and faster computing, accurately scoring content quality in a generic scoring model is a more distal goal.

Writing Pal

Writing Pal is a web-based tutoring system that provides writing strategy instruction and practice (Crossley, Allen, & McNamara, in press; Roscoe & McNamara, 2013). High school students are the primary audience for Writing Pal, but it can be used with students in middle school through college. Writing Pal focuses on persuasive-style essays, although many of the strategies apply to other types of writing as well. The system provides lesson videos across nine topics that span the entire writing process, from prewriting to drafting to revising. Each lesson video is presented by a pedagogical agent and covers a specific strategy. For example, in the Conclusion Building topic, one of the videos provides strategies to maintain readers' interest with a strong closing. Checkpoint questions at the end of videos provide feedback to students about how well they understood the lesson content. For each topic, there are practice games that reinforce the strategies taught in the lessons. Some activities include generative practice in which students draft writing samples and receive feedback on their performance.

In addition to strategy instruction and practice games, students can write and revise entire essays in the Writing Pal system and receive an automated score and feedback. The automated essay scoring is driven by an algorithm that is powered by several NLP tools (see McNamara et al., 2013). For example, measures of word sophistication and text cohesion are calculated from linguistic indices and are included in the scoring algorithm. Similar to Criterion, the algorithm is built using expert-scored essays as the scoring benchmark. Writing Pal uses a generic model for scoring, allowing the same algorithm to be used for many prompts, including prompts entered by teachers. Separate algorithms drive the selection of feedback for each essay. First, the essay is checked for length; if it fails, students will receive feedback on content generation. Next, the system checks for structural elements of the essay. For example, if an essay comprises two long paragraphs, feedback will focus on how to structure an essay into an Introduction, Body, and Conclusion. If these initial checks are passed, the introduction, body, and conclusion paragraphs are individually assessed using different algorithms. These paragraph-level algorithms use linguistic indices to make inferences about paragraph quality. For example, if the conclusion to an essay is flagged as being of lower quality, the student will receive strategy feedback on how to improve a conclusion.

Importantly, all feedback messages within Writing Pal are actionable and reference a strategy taught in the Writing Pal system. The alignment between instructional content and feedback messages ensures that students can seek help on topics that are difficult for them. The feedback promotes higher-level strategy use and is not specific to any one prompt, which makes it possible for teachers to insert their own essay prompts. The flexibility afforded by the NLP tools that drive the scoring and feedback allows teachers to easily insert new content in the system, focusing students on curriculum-relevant practice.

Scaffolded Writing and Rewriting in the Discipline (SWoRD)

SWoRD is a system that supports peer review in the classroom and is designed for high school and college students (Schunn, in press; Patchan, Hawk, Stevens, & Schunn, 2013). Because writing practice is vital for developing writers, peer review can allow students to receive feedback on more writing assignments than a single teacher could provide. In addition to providing students with more feedback on their work, the process of writing feedback is also beneficial for students (Cho & MacArthur, 2011). Thus, instead of attempting to provide automated feedback to students (e.g., Writing Pal and Criterion), SWoRD provides a platform for teachers to set up peer review among students.

SWoRD provides several supports for peer review. The system uses a web-interface to anonymously assign students with papers that they review, and then returns reviewed papers to the original author. Each paper is assigned to several reviewers (the exact number being defined by the teacher), and students can elect to review additional papers to receive bonus points. A reviewing form is provided to reviewers for each assignment. This form is customized by the teacher (including the dimensions on which to rate the essay) and includes examples of the types of comments that students should include. After authors receive their reviews, they rate them on their helpfulness (called "back-evaluations). Reviewers' ratings are compared to the averages for each essay they review and a "reviewer accuracy" score is calculated.

The SWoRD system represents an elegant solution to the need for additional feedback on writing assignments beyond what a teacher can reasonably provide to each student. Although asking students to provide feedback does require more work from them than automated scoring, the task of peer editing is beneficial. Moreover, peer review sidesteps some of the major challenges with automated scoring, such as difficulties with providing accurate feedback on content correctness for a wide array of essay prompts. SWoRD also helps address common problems with peer review, such as a lack of effort on the part of the reviewer or a tendency to be overly positive (VanDeWeghe, 2004). First, it anonymizes the review process. Second, by calculating helpfulness (through the back-feedback) and reviewer accuracy scores, students are presented with information about how they are doing *as a reviewer*, which helps keep them accountable. Overall, SWoRD's success is an important reminder that systems can support literacy through tools that enable instantiations of existing pedagogical methods.

Research Writing Tutor (RWT)

RWT was designed to provide writing instruction to graduate students, specifically for research writing (Cotos, in press). Although graduate students have already achieved academic success, they have not necessarily mastered the skills required for composing compelling, thorough research articles. Formal instruction in this topic is rare and less well studied than many other areas of literacy. RWT provides scaffolded feedback to undergraduate and graduate students who are writing research articles. The system uses a corpus of domain-specific articles that provide examples of how to achieve the goals of scientific writing; these articles also feed into the feedback system. The pedagogy behind the system is based on the analysis of 900+ articles across 30 domains (Cotos, Huffman, & Link, 2015). More specifically, RWT focuses on *moves*, which are communicative goals, and *steps*, which are the strategies used to achieve the moves. Cotos and colleagues (2015) identified a set of moves and steps for the Introduction, Method, Results, and Discussion sections. The moves and steps were defined based on the English for Academic Purposes (EAP) field grounded in Swale's (1981) genre theory. As an example, the first move in the Method section is "contextualizing the study method" and the first step within this move is "referencing previous works." Each of the research articles included in the RWT was tagged for each of the moves and steps that were included in any given article.

Three main components make up RWT: learning, demonstration, and feedback. In the learning module, students are instructed on the conventions of scientific writing specific to their domain. The content in this module provides explanations for the moves and steps, including short videos in which an instructor describes the moves and steps used in excerpts from the corpora. The lesson module also provides lessons on language use in academic writing, such as the type of language used to draw comparisons. In the demonstration module, students are presented with complete research articles from the corpora. They can then view examples of individual sections (e.g., Introductions) that are annotated with moves and steps. Different colors represent segments of the sections that belong to one of the moves, and mousing over a particular section will display the step that it represents. The demonstration module also provides students with a search engine in which students can find examples of particular steps, potentially narrowing their search by genre. By viewing multiple examples of the same step, students can see how multiple authors achieve similar goals. Finally, in the feedback module, students can input their own scientific writing and observe information about which moves and steps the system has identified in their paper. RWT assigns moves and steps to students' writing using Support Vector Machine classifiers trained using the hand-annotated corpora (Cotos & Pendar, in press). The feedback provides both macro-level (i.e., at the move level) and micro-level (i.e., at the steps level) feedback that can help students identify strengths and weaknesses of their writing. As part of this feedback, students are asked clarifying questions about their writing, and are provided with relevant links to the lesson and demonstration content.

The team behind RWT has met the considerable challenge of providing personalized feedback to advanced scientific writing. They did so by first hand-annotating a large corpus using a theoretically and pedagogically supported framework, and then building a classifier for students' writing using these annotations. A limitation to this approach is that adding new domains is time consuming, and attempting to cover *all* scientific topics is unrealistic. However, the system as-is already provides instruction and feedback that is generally useful, even across domains that were not represented in the training corpora.

Future Technologies

In this chapter, we provided overviews of eight literacy technologies, each of which has found success despite the challenges of developing technology for an ill-defined domain such as literacy. Some of these systems have expert-annotations for a large database of materials, allowing useful, personalized feedback to be delivered to students (e.g., ITSS, RWT). Many of the systems use various NLP techniques to adapt to students' written responses (e.g., DSCoVAR, iSTART-2, Criterion, Writing Pal). Thus, the *intelligence* in NLP-driven ITSs can be achieved through analyses of student language. As NLP techniques improve and become more widely used and developed, these types of systems are expected to garner capabilities to more accurately assess student-generated content and provide more adaptive feedback and training.

Tools are already emerging that help provide access to sophisticated NLP techniques to non-experts (also see Crossley, Allen, & McNamara, 2014). For example, The Coh-Metrix Common Core Text Ease and Readability Assessor (TERA; Jackson, Allen, & McNamara, in press) allows teachers, researchers, and even students to enter texts to be analyzed by an NLP engine on several dimensions such as narrativity and syntactic simplicity. An ITS developer may use the output from a system such as TERA to help adaptively select texts for students at different grade levels. An enduring challenge, both intellectually and practically, is the ability to develop flexible and accurate algorithms using NLP. Many algorithms apply only to a specific type of language input and are built to predict a relatively small number of measures (e.g., essay score). Larger, well-annotated corpora and *big data* may open doors to building NLP tools and algorithms that provide a broader range and more accurate information about language and performance.

Future technologies are also expected to extract and use a wider array of information. Most NLP algorithms use only the information that is in the language input (e.g., essay) to predict an outcome, such as a student's writing ability. Adaptive technologies may also incorporate information about students' prior abilities, such as their prior reading or writing ability (Crossley, Allen, Snow, & McNamara, 2015). In turn, current work is being conducted to circumvent the need for the multiple assessments that are needed to measure prior abilities. For example, information about students' vocabulary knowledge (Allen & McNamara, 2015), reading ability (Allen, Snow, & McNamara, 2015; Varner, Jackson, Snow, & McNamara, 2013), affective states (Allen et al., in press), and cognitive processes (Allen, McNamara, & McCrudden, 2015) can be extracted from linguistic features of students' written responses. By using NLP to estimate students' prior abilities, interests, and motivation levels as well as their performance within the system, better student models will emerge, with fewer assessments.

While the quality of the student model is important to ITSs, these systems are often intended to be used in classrooms, and generally by teachers. Another theme that emerged across these systems is the need for tools that are provided to the teachers who use them (e.g., A2i, SWoRD, iSTART-2, Criterion, Writing Pal). A2i is a system directed specifically at teachers, providing suggestions for personalizing comprehension instruction to students. Its development has had a strong focus on usability for teachers, which is crucial for its adoption. SWoRD also places primary importance on its interface for both teachers and students. This system provides a tool for teachers to enable peer review, providing the ability for teachers to set the specific dimensions on which students rate their peers. iSTART-2, Criterion and Writing Pal both provide means for teachers to add custom content (texts that can be self-explained in iSTART-2 and essay prompts in Criterion and Writing Pal). NLP-driven algorithms that rely on generic models allow feedback to be provided for teacher-authored system content.

The Language Muse Activity Palette is another system that is currently under-development that also employs NLP for the direct benefit of teachers and their ability to create course content for English language learners in middle school (Burstein & Sabatini, in press). The Palette analyzes texts entered by teachers and creates activities based on the text content. There are a range of activities, including word-level activities (e.g., questions about content words) and sentence-level activities (e.g., questions about how words signal relations between parts of a sentence). The suite of activities that teachers create within Palette will allow individualized practice that addresses students' individual needs and teachers' particular curricular requirements.

In sum, educational technologies for literacy instruction have found success despite difficult challenges. They do so by taking different approaches: by using NLP techniques to score individual responses, by annotating large corpora of materials, and by facilitating existing pedagogical methodologies. In the future, systems will move toward greater adaptivity to individual students, likely using combinations of these methods. In addition to improving adaptivity, developers will continue to make the systems more usable for teachers in specific educational contexts. As the field considers both the technology and how it fits into the classroom, literacy instruction will adapt to the considerable instructional needs of students from early development to advanced technical reading and writing.

References

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)*. Madrid, Spain.

Allen, L.K., McNamara, D.S., & McCrudden, M.T. (2015). Change your mind: Investigating the effects of self-explanation in the resolution of misconceptions. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio, *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (Cog Sci 2015)*. Pasadena, CA.

Allen, L.K., Mills, C., Crossley, S.A., Jacovina, M.E., D'Mello, S.K., & McNamara, D.S. (in press). Investigating boredom and engagement during writing using multiple sources of information: The essay, the writer, and keystrokes. To appear in the *Proceedings of the 6th International Learning Analytics & Knowledge Conference (LAK'16)*.

Allen, L.K., Snow, E.L., & McNamara, D.S. (2015). Are you reading my mind? Modeling students' reading comprehension skills with Natural Language Processing techniques. In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron, & G. Siemens (Eds.), *Proceedings of the 5th International Learning Analytics & Knowledge Conference (LAK'15)* (pp. 246-254). Poughkeepsie, NY.

Boonthum-Denecke, C., Levinstein, I.B., McNamara, D.S., Magliano, J.P., & Millis, K.K. (2008). NLP techniques for intelligent tutoring systems. In J. Rabuñal, J. Dorado & A. Pazos (Eds), *Encyclopedia of Artificial Intelligence* (pp. 1253-1258). Hershey, PA: Idea Group, Inc.

Burstein, J., & Sabatini, J. (in press). The *Language Muse Activity Palette*: Technology for promoting improved content comprehension for English language learners. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Cheung, A.C., & Slavin, R.E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis.*Educational Research Review*, *7*, 198-215.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*, 73-84.

Collins-Thompson, K., & Callan, J. (2007). Automatic and human scoring of word definition responses. *Proceedings of the North American Association of Computational Linguistics-Human Language Technology*, (pp. 476-483).

Connor, C.M., Morrison, F.J., Fishman, B., Crowe, E.C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, *24*, 1408-1419.

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

Cotos, E. (in press). Computer-assisted research writing in the disciplines. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes, 19,* 52-72.

Cotos, E., & Pendar, N. (in press). Discourse classification into rhetorical functions for AWE feedback. *CALICO Journal*, *33*.

Council of Writing Program Administrators, National Council of Teachers of English, & the National Writing Project. (2011). *Framework for success in postsecondary writing.* Retrieved from http://wpacouncil.org/files/framework-for-success-postsecondary-writing.pdf

Crossley, S.A., Allen, L.K., & McNamara, D.S. (in press). The Writing Pal: A writing strategy tutor. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Crossley, S.A., Allen, L.K., & McNamara, D.S. (2014). Analyzing discourse processing using a simple natural language processing tool (SiNLP). *Discourse Processes, 51*, 511-534.

Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst...Textual features...There is more to Automatic Essay Scoring than just you!. In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron, & G. Siemens (Eds.), *Proceedings of the 5th International Learning Analytics & Knowledge Conference (LAK'15)* (pp. 203-207). Poughkeepsie, NY.

Crossley, S.A., & McNamara, D.S. (Eds.). (In press). *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*, 7-24.

Desmarais, M.C., & Baker, R S.J.D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*, 9–38.

Ericsson, K.A. (2008). *Deliberate practice and acquisition of expert performance: A general overview.* Academic Emergency Medicine, 15, 988–994.

Ericsson, K.A., Krampe, R.T., & Tesch-Römer, C. (1993). *The role of deliberate practice in the acquisition of expert performance.* Psychological Review, 100, 363–406.

Fournier-Viger, P., Nkambou, R., & Mephu Nguifo, E. (2010). Building intelligent tutoring systems for ill-defined domains. In Nkambou, R., Mizoguchi, R., Bourdeau, J. (Eds.). *Advances in Intelligent Tutoring Systems*, Springer, p.81-101.

Frishkoff, G.A., Collins-Thompson, K., Nam, S., Hodges, L., & Crossley, S. (in press). Dynamic Support of Contextual Vocabulary Acquisition for Reading (DSCoVAR): An intelligent tutor for contextual word learning. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Glenberg, A.M., Walker, E.A., & Adelaida Restrepo, M. (in press). EMBRACEing Dual language learners. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Ingebrand, S.W., & Connor, C.M. (in press). Assessment-to-Instruction (A2i) – An online platform for supporting individualized early literacy instruction. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Jackson, G.T., Allen, L.K, & McNamara, D.S. (in press). Common Core TERA: Text Ease and Readability Assessor. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Jackson, G.T., Guess, R.H., & McNamara, D.S. (2010). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science, 2,* 127–137.

Jackson, G.T., & McNamara, D.S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105,* 1036-1049.

Kegel, C.A., & Bus, A.G. (2012). Online tutoring as a pivotal quality of web-based early literacy programs. *Journal of Educational Psychology, 104,* 182.

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

Le, N.T., Loll, F., & Pinkwart, N. (2013). Operationalizing the continuum between well-defined and ill-defined problems for educational technology. *IEEE Transactions on Learning Technologies*, *6*, 258-270.

Leahy, S., Leusner, D., Lyon, C., King, T., Ridolfi, L. (2014). *Product Efficacy Argument for the Criterion® Online Writing Evaluation Service Third Edition.* PEAr-14-01. Princeton, NJ: Educational Testing Service.

Lynch, C., Ashley, K.D., Pinkwart, N., & Aleven, V. (2009). Concepts, structures, and goals: Redefining ill-definedness. *International Journal of Artificial Intelligence in Education*, *19*, 253-266.

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

McNamara, D.S., Crossley, S.A., & Roscoe, R.D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods, 45,* 499-515.

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. & Magliano, J.P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation.* New York, NY: Elsevier Science.

Meyer, B.J.F., & Wijekumar, K. (in press). Intelligent Tutoring of the Structure Strategy: A reading strategy tutor. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Meyer, B.J.F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly, 45*, 62-92.

Meyer, B.J.F., Wijekumar, K.K., Lin, Y. C. (2011). Individualizing a web-based structure strategy intervention for fifth graders' comprehension of nonfiction. *Journal of Educational Psychology, 103*, 140-168.

Neuwirth, M. (2014). Issues in Intelligent Tutoring Systems : Exploring issues in learning and teaching writing, *Computers and the Humanities*, *23*, 45–57.

Patchan, M.M., Hawk, B., Stevens, C.A., & Schunn, C.D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*, 381-405.

Ramineni, C., & Deane, P. (in press). The *Criterion*® Online Writing Evaluation Service. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Ramineni, C. & Williamson, D.M. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing*, *18*, 40–61.

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*, 1010-1025.

Schunn, C. (in press). Writing to learn and learning to write through SWoRD. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Snow, E.L., Jacovina, M.E., Jackson, G.T., & McNamara, D.S. (in press). iSTART-2: A Reading Comprehension and Strategy Instruction Tutor. To appear in Crossley, S.A. & McNamara, D.S. (eds.), *Adaptive Educational Technologies for Literacy Instruction.* Taylor & Francis, Routledge: NY.

Swales, J.M. (1981). *Aspects of articles introductions.* Aston ESP Reports, No. 1. The University of Aston in Birmingham.

VanDeWeghe, R. (2004). Awesome, dude! Responding helpfully to peer writing. *English Journal, 94*, 95-99.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*, 227-265.

Varner (Allen), L.K., Jackson, G.T., Snow, E.L., & McNamara, D.S. (2013). Linguistic content analysis as a tool for improving adaptive instruction. In K. Yacef et al. (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED),* (pp. 692-695). Heidelberg, Berlin: Springer.