Intelligent Reading Tutor for Improving Reading Comprehension and Content Learning

Jacqueline Haynes, Ph.D.
Intelligent Automation, Inc.
Rockville, MD
jhaynes@i-a-i.co

Daniel Fowler
Intelligent Automation, Inc.
Rockville, MD
dfowler@i-a-i.com

Shannon Mayhew
Intelligent Automation, Inc.
Rockville, MD
smayhew@i-a-i.com

ABSTRACT

Reading comprehension is a critical skill for job performance and advancement within the military. Many newly enlisted and veteran soldiers experience difficulties in mastering core content to meet basic and advanced requirements because of poor comprehension skills. For example, it has been noted that twenty-five percent of the Navy's enlisted population, score below eighth grade levels in basic academic skills. These failings delay assumption of duty and limit job advancement, both of which are detrimental to performance and efficiency in the military. GradAtions, a web-based intelligent tutor funded through the Navy, helps in the improvement of reading comprehension skills and the ability to learn new information from text in an efficient manner. It may be used with any text including technical manuals, workplace related documents, instructional materials, etc. thus supporting content learning while developing comprehension. Based on well recognized cognitive models in reading and learning, the technology utilizes summarization to improve comprehension and to measure how well a learner understands the information. The key innovation in this intelligent tutor is the design and integration of several text analysis algorithms, including semantic analysis, for performing the automated evaluation of the natural language summaries provided by the learners. The design also recognizes the strong correlation between vocabulary and comprehension. Its "tutors" assist with vocabulary, sequencing, fluency as well as summarization. Utilization of GradAtions in three different studies has consistently demonstrated improvement in reading comprehension skills of one year or more. The most recent study at Great Lakes Naval Training Center has noted preliminary results of over 2 grade levels.

ABOUT THE AUTHORS

Jacqueline Haynes received her Ph.D., B.A., and M.Ed. from the University of Maryland. She is Founder and Executive Vice President of Intelligent Automation, Inc. (IAI). Dr. Haynes' background is in learning and instruction, especially using artificial intelligence and other advanced technologies. Her R&D efforts for the past 20 years have focused on adaptive instruction, reading comprehension, and on tools to support high quality instruction within a distributed distance learning environment. She was previously a classroom teacher, a reading supervisor for a large school district and a faculty member at the University of Maryland, Department of Special Education and a consultant in artificial intelligence.

Daniel Fowler is a research engineer at Intelligent Automation, Inc. Mr. Fowler has an academic background in Artificial Intelligence/Cognitive Science, and has since worked as both a developer and an evaluator of educational technology. He is interested in developing instructional software that embodies educational and psychological theories, and that take advantage of educational standards such as SCORM. At IAI, he has served as technical lead for the development of several tools to support distributed distance learning.

Shannon Mayhew is a Senior Instructional Designer at Intelligent Automation, Inc. She received her M.Ed. from Harvard University, developed instructional materials for well-known children's books and television (including Sesame Street), and currently works on transitioning effective instructional methods from the classroom to the computer.

Intelligent Reading Tutor for Improving Reading Comprehension and Content Learning

Jacqueline Haynes, Ph.D.
Intelligent Automation, Inc.
Rockville, MD
jhaynes@i-a-i.co

Daniel Fowler
Intelligent Automation, Inc.
Rockville, MD
dfowler@i-a-i.com

Shannon Mayhew
Intelligent Automation, Inc.
Rockville, MD
smayhew@i-a-i.com

INTRODUCTION

Highly advanced interactive computer-based training courses for warfighters are being developed as technology plays an increasingly significant role in their training. Job aids and other field operation guides also contribute to warfighter performance enhancement. But these programs are only as effective as the readiness of the warfighters to benefit from them. Lack of certain basic skills, like reading comprehension, can significantly hinder the effectiveness of training and performance support systems. This problem is magnified by increased reliance on computer-based training since such systems still rely quite heavily on text for conveying information. Having a minimum standard for recruitment often does not guarantee that recruits will possess adequate reading comprehension skills. Twenty-five percent of the Navy's enlisted population, for example, scores below the eighth grade level in reading, writing, and arithmetic. Beyond the military, literacy is one of the most fundamental requirements for succeeding in today's world. Yet, 50 million Americans are not functionally literate. With the growth in availability of enormous amounts of free information via the Internet and the Web, the plight of people with limited literacy can only worsen. They will be unable to take full advantage of this new resource and fall further behind in managing their lives as access to the large amounts of public information becomes an accepted norm. Given this background, the Navy undertook an effort to support research and development in computer-based programs to help Navy enlisted personnel with limited literacy to improve their ability to read and understand Navy information presented to them as text.

The present paper is the result of an Office of Naval Research-sponsored project whose goal was to create instructional software to build more proficient reading skills within Naval and other service personnel. A solution to this challenge starts with identifying what skills make someone a good reader. Good readers are active readers. From the outset they have clear goals in mind for their reading. They constantly evaluate whether the text, and their reading of it, is meeting their goals. Good readers typically look over the text before they read, noting such things as the structure of the text and text sections that might be most relevant to their reading goals. As they read, good readers frequently make predictions about what is to come. They read selectively, continually making decisions about their reading--what to read carefully, what to read quickly, what not to read, what to re-read, and so on. Good readers construct, revise, and question the meanings they make as they read. They draw upon, compare, and integrate their prior knowledge with material in the text. They think about the authors of the text, their style, beliefs, intentions, historical milieu, and so on. They monitor their understanding of the text, making adjustments in their reading as necessary. Good readers try to determine the meaning of unfamiliar words and concepts in the text, and deal with inconsistencies or gaps as needed. They evaluate the text's quality and value, and react to the text in a range of ways, both intellectual and emotional. Good readers read different kinds of text differently. For example, when reading narrative, good readers attend closely to the setting and characters; when reading expository text these readers frequently construct and revise summaries of what they have read. For good readers, text processing occurs not only during 'reading' as we have traditionally defined it, but also during short breaks taken during reading, and even after the 'reading' itself has commenced. Comprehension is a consuming and complex activity, but one that, for good readers, is typically both satisfying and productive (Duke & Pearson, 1998).

Building good reading skills in adults

Given knowledge about what good readers do when they read, researchers and educators have addressed the question: Can we teach students to engage in these productive behaviors? The answer to this question is a resounding yes! A large volume of work indicates that we can help students acquire the strategies and processes used by good readers–and that this improves their overall comprehension of text, both the texts used

to teach the strategies and texts they read on their own in the future. Building good reading skills consists of the following:

- A great deal of time spent actually reading: As with decoding, all the explicit instruction in the world will not make strong readers unless accompanied by lots of experience applying their knowledge, skills, and strategies during actual reading.
- Experience reading real texts for real reasons. In order to become strong, flexible, and devoted comprehenders of text, students need experience reading texts beyond those designed solely for reading instruction, and experience reading text with a clear and compelling purpose in mind.
- Experience reading at least the range of text genres that we wish students to comprehend. Students will not learn to become excellent comprehenders of any given type of text without substantial experience reading and writing it. Thus, for example, all the experience in the world reading storybooks will not, by itself, enable a student to read, understand, and critique procedural forms of text of the sort found in how to books, instructions manuals, and the like.
- An environment rich in vocabulary and concept development, through reading, experience, and, above all, discussion of words and their meanings. Any text comprehension depends on some relevant prior knowledge. To some degree, well-chosen texts can, in themselves, build readers' knowledge base. At the same time, hands-on activities, excursions, conversations, and other experiences are also needed to develop vocabulary and concept knowledge required to understand a given text.
- Substantial facility in the accurate and automatic decoding of words. In a recent review of the literature, Pressley (in press) argues compellingly that skilled decoding is necessary, though by no means sufficient, for skilled comprehension.
- Lots of time spent writing texts for others to comprehend. Again, students should experience writing the range of genres we wish them to be able to comprehend. Their instruction should emphasize connections between reading and writing, developing students' abilities to write like a reader and read like a writer. (Duke & Pearson, 1998)

With any training requirement, there are always multiple approaches to achieving training goals. While reading has traditionally been taught in classroom settings, over the years technology has offered some cost-effective, convenient and scalable alternatives. The remainder of the paper describes an innovative, web-based- approach to teaching reading comprehension to adults that embodies the reading instruction principles described above. We begin by describing traditional software-based approaches then describe GradAtions, IAI's approach to reading instruction.

Teaching Reading in the Navy

Computer-based training software for reading tends to fall into one of two categories, both of which are inappropriate for the Navy's target group. Commercially-available adult literacy software tends to target truly illiterate adults - those who cannot read at all. These programs tend to focus on a lower level of literacy than is present among Navy recruits, and typically is focused on developing phonics and word decoding skills and building sight vocabulary. However the Navy requires a basic ability to read for enlistees. Navy personnel have mastered the skill of decoding, at least at its most basic level.

Another type of software addresses reading comprehension skills, but has content that is typically geared toward children, thus inappropriate for adults. It is critical that programs to be used by adults use adult-relevant content. Furthermore, for the Navy, it is desirable that the content include material relevant to Navy personnel, such as technical manuals, general materials describing Navy procedures and policies, documents describing general citizenship, etc. Also appropriate would be material, such as life-skills (i.e., as balancing a checkbook, understanding instructions for taking medication), descriptions of regions, countries, and cities with which the Navy is currently involved, news, parenting tips, etc.

Another important requirement for the Navy is that personnel should be able to train themselves through independent study. There are no reading instructors aboard ships where enlisted personnel spend much of their time. Therefore, the software should not require an instructor or classroom.

To address these problems, the Navy funded research on the development of GradAtions ®, an ITS that enhances reading comprehension by focusing on summarization for continuous assessment and skill development. We will also discuss the instructional design, technical approach and the results of an evaluation studies on the effectiveness of this approach.

GRADATIONS: A READING COMPREHENSION TUTOR FOR ADULTS

GradAtions is a Web-based reading comprehension tutor that uses *any text* to develop reading comprehension skills among adults who have already acquired basic decoding skills, but do not have sufficient skills to read and understand materials required for their jobs. GradAtions, therefore, does not include any particular content or domain information, but rather provides the ability for the customer either to provide or to specify the text to be used. This text could be of four types: (1) workplace-related text such as training manuals, technical manuals, employee manuals, in-house training materials; (2) academic text such as textbooks or other instructional materials; (3) public information text such as information related to parenting skills, nutrition and health, public safety; and/or (4) high-interest text to encourage motivation to read and improve skills such as magazine text (i.e., sports, gardening, travel), or material from popular sites on the Web. The design for GradAtions is based on the concept of "reading ladders," where a reader begins reading materials within a domain that captures his/her interest and the specific text is easily understood. Subsequent passages are selected within the same domain at gradually increasing difficulty levels, allowing the reader to gain background knowledge and vocabulary, and to avoid reading experiences that are frustrating by being too difficult.

GradAtions™ Instructional Approach

GradAtions uses robust, in-depth multiple assessments of reading comprehension and automated tutoring and guidance techniques, to help adult learners improve their reading and comprehension skills as they are learning meaningful and relevant subject matter.

The reason we named our approach GradAtions is that the system presents an individually adaptive, dynamically arranged progression of reading texts within a subject matter domain, and uses an automated guidance tool to match individual students with text that best suits their current reading competence and their "zone of proximal development" (Vygotsky 1978), based on the match between the text's content and the reader's current knowledge of the domain. The system evaluates the difficulty level of new materials for an individual reader, presents the reader with material at an appropriate level of difficulty, and then guides the reader through what we visualize as a "ladder" structure defined by the relative difficulty and relatedness of portions of the reading material. While the student is always in control of what he or she will read next, the automated guidance tool makes recommendations that are designed to keep the student on a course that is appropriately challenging: not too difficult and not too easy. This process is illustrated below.

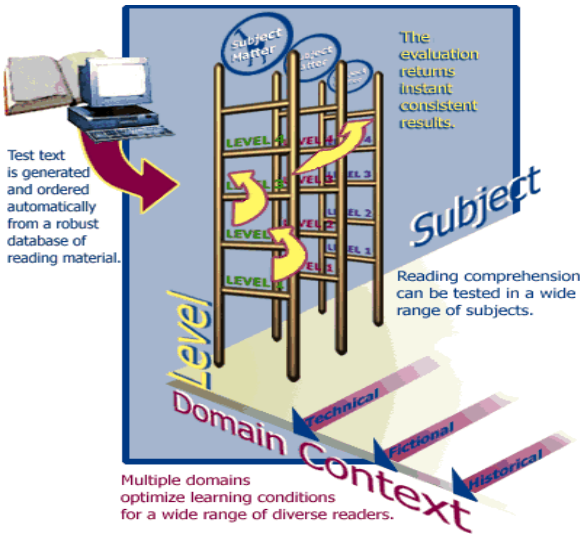

Figure 1. The GradAtions Model of Reading and Development

In this way, GradAtions attempts to optimize both content learning and improvement of reading comprehension skills. The tutors for vocabulary, comprehension and summarizing facilitate this progression. GradAtions summarizes both as a strategy for improving the student's ability to produce a good summary, and as a central method of assessing comprehension. According to Brown and Day (1983), when learners write summaries, they create "macrorules," or the mental operations used by expert readers when they remember the gist of what they are reading. This is a crucial skill to develop reading comprehension and to learn new information from text. Brown and Day (1983) studied summarization abilities in fifth grade, seventh grade and college level students. They concluded that summarization is not easily mastered at any age. They determined that fifth grade students routinely applied only the first two rules, while seventh graders and college students employed the first four. College used the fifth rule students in only 50 percent of the appropriate cases. Brown and Day (1983) go on to explain that the two younger groups tend to quote and delete sections from the original, whereas older students are more adept at condensing and rephrasing.

Hare and Borchardt (1984) found that students who do not have enough background knowledge to determine

what is important in a text have difficulties with summarization. Additional problems occur when students are not interested in a topic, or have poor writing abilities (Head and Buss, 1987).

Summarizing Instruction

Many studies have tested the effectiveness of teaching summarization, as a means of increasing textual understanding, as measured by knowledge retention.

Research prior to 1980 (e.g. Arnold, 1942; Germane, 1921) generally failed to find specific advantages in summarization instruction over unstructured reading. These results stand in startling contrast to those of studies conducted after 1980, which have shown considerable improvement in comprehension. Armbruster and Anderson (1984) attribute this to differences in methodology: the former generally used multiple choice assessment, which is unlikely to reflect the improvements achieved by summarization. In contrast, later research used open-ended questioning to test recall, which is more likely to reveal a difference.

One such study is described by Taylor (et al, 1985). In this study, summarization strategy is modeled by a teacher, using a think-aloud procedure, then practiced by students, initially with the teacher's guidance, and then independently. Using this method, they found greater improvements in text recall than with two other strategies (question-answering and additional studying). Unfortunately, improvements were only found when testing material from the same topic.

Bean and Steenwyk (1984) found an improvement in summarization skill after a short number of training periods (12x 30 minutes). They also found associated improvements in standardized reading test scores. Similarly Hare and Borchardt (1984), noted that students could be trained to extract the main idea of a text if it was explicit, but had great difficulty if it was implicit. Rinehart (et al 1986) found that summarization produced improvement in recall of main ideas in the text, but not minor details.

Gude, Jackson and Shaw (2000) describe interventions that targeted first, fifth, and eighth grade students with low standardized test scores. These students were taught summarization, in addition to other pre- and post- reading strategies. They report consistent improvements in reading comprehension ability. Aarnoutse, Brand-Gruwel and Oduber (1997) used clarifying, questioning, summarizing and predicting with a different modality, listening. They found improvements in reading, but not listening comprehension.

The GradAtions system measures the semantic similarity between a learner's passage summary and the underlying semantic structure of the original text. Natural language analysis algorithms are used to evaluate text difficulty levels for an individual student among a variety of semantically related reading passages. The system provides feedback to the user on portions of important concepts in the text that his/her summary is missing. Furthermore, information about concepts that are included or not included in a learner's summary is then used by GradAtions to suggest the next passage in the "ladder" that the learner should read. It is also used to suggest particular tutor activities that will help the learner improve his particular area of weakness in understanding a specific text. Scaffolded activities such as these have been proven as effective methods for literacy instruction (Palincsar and Brown, 1984).

Vocabulary Instruction

From a methodological perspective, the most important advantage of vocabulary instruction is that, like summarization, it also has an effect upon reading comprehension. Therefore, the effects of vocabulary instruction can be compared on the same dependent measures of comprehension. One example of this effect is in Beck et al. (1982), where fourth-grade children were taught 104 new vocabulary words over five months. Students receiving this instruction demonstrated better comprehension than control students.

Several studies illustrate the link between the extent of a reader's vocabulary and comprehension skills (Anderson & Freebody, 1981; Nagy, Anderson & Herman, 1987; National Reading Panel, 2000). More importantly, many experiments indicate that an extensive vocabulary actually promotes skilled comprehension (Beck, Perfetti & McKeown, 1982; McKeown, Beck, Omanson & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985).

In seeking reasons behind this link, researchers report conflicting findings. Some studies suggest that vocabulary words taught in context improved comprehension more significantly than words taught in isolation (Beck, Perfetti & McKeown, 1982; Blanchard, 1980; Maberry, 2000; National Reading Panel, 2000). Yet other studies report instruction where vocabulary words taught in isolation, with more emphasis on rapid decoding, was more effective (Breznitz, 1997a; Breznitz, 1997b; Peitz and Vena, 1996; Tan and Nicholson, 1997). The former argument may seem the more logical, as in this instructional method, readers are actively processing words within their context. However, one theory for the results in

the latter argument is that repeated practice in decoding results in rapid recognition, thus "freeing up more short-term capacity for comprehension" (Nagy and Scott, 2000).

Content Tutoring

While the Navy's primary purpose for GradAtions is to improve reading comprehension skills, a correlated benefit lies in improving the reader's knowledge of the domain. Since GradAtions can include any text at any level of difficulty (above approximately 4th grade level), a reader who does not have reading difficulty could use GradAtions solely for the purpose of learning the information presented in the text. Hence, GradAtions can be viewed as a content tutor as well. In this context, if the material were, for example, a textbook used in a frequently-taken course in a community college, students taking the course could use GradAtions to help them to understand the material or for self-assessment to help guide their study of the textbook. Students whose reading skills were inadequate or marginal for the textbook (as is true for nearly 25% of community college students) would get help and support from the Gradations environment in understanding the text.

Instructional Method and Assessment

Instruction in GradAtions is based on the principle that the best way to improve reading skill is by reading. This principle dictates that more instructional time be spent reading than in doing related activities that do not involve much reading. A second principle is that comprehension improves when instruction takes place at a level that is within an individual reader's "zone of proximal development," which means that the material includes a significant amount of information and vocabulary the reader already knows, as well as new information that the reader can understand and conceptually organize in his/her individual knowledge base. A third principle is based on a large body of research showing that summarizing is an excellent method of learning to comprehend a passage (primarily by focusing on important information, deep processing associated with relating the concepts in the passage to prior knowledge and associated vocabulary) and that the quality of a summary is also a good measure of how well a reader has understood a passage. Finally, the design recognizes the strong correlation between vocabulary and comprehension.

Instruction and assessment in GradAtions work in the following manner: A reader selects an appropriate passage to read, and reads it. At any time during reading, he/she can use any of the several GradAtions 'tutors' to assist with vocabulary, sequencing, or summarizing. When ready, he/she then writes a summary of the passage. The summary can be written either independently, or within a structured, supported environment for summary writing called "Summary Shop." The completed summary is submitted to GradAtions for scoring, and the reader then (typically with about a 10 second delay) gets feedback on the summary and is directed to read the next passage (for a good summary), improve the summary (for a summary that shows partial comprehension), or to re-read the passage using the tutors to develop a better understanding of the passage (for a poor summary). At any time the reader can move on to another passage, regardless of whether or not he/she has written a good summary for the current passage. GradAtions selects the next passage for an individual reader based on his/her specific performance with the current passage and its relationship (in terms of readability and content) to other passages. These features make GradAtions much like an intelligent tutoring system (ITS), without many of the technical obstacles and cost associated with ITS development. Our experience is that on average, it takes approximately 40 minutes for a reader to read a passage, use the tutors, and write a summary for a single passage. This estimate assumes that the passage is well-matched to the reader's current abilities and background knowledge.

Technology behind GradAtions

Summary scoring is a key technical feature of GradAtions and a technological innovation that differentiates GradAtions from other technology-based approaches to reading comprehension improvement. Gradations includes a set of scoring algorithms based on statistical methods of natural language processing (IAI's proprietary algorithm is called HiWISE™), as well as algorithms that detect a copy/delete method (the copy and paste capability are disabled), summaries that are too long or too short, and other measures of summary quality. Scoring protocols follow those used in highly regarded summarizing research. Studies of the HiWISE algorithm, show that on passages consisting of more than 20 words, HiWISE scores summaries, as well as teachers who have received considerable training in summary scoring.

Using statistical semantic methods to evaluate summaries

Statistical methods vs. Traditional A.I.

One of the key features of statistical (a.k.a. empirical) models of language is that they achieve human

correspondence without using domain specific hand coded knowledge and inference rules. Instead, they apply nothing more than unsupervised learning algorithms to training material (corpus). The text in a corpus is commonly pre-processed, for example by lemmatization, which means replacing inflected or variant word forms (e.g. 'spoke', 'speaks', 'speaking') with their base form, which in this case is 'speak'. The words of a corpus can also be tagged for syntactic 'part of speech', verb, noun, adjective, etc. Various statistics can then be derived from analyzing a corpus, including individual word frequencies, co-occurrence data (the frequency that one word occurs in proximity to another), and contextual distinctiveness (the degree to which a word is used specifically in certain contexts). These statistics numerically characterize aspects of general word usage, and can be applied to model certain cognitive/linguistic abilities on a computer.

HiWISE™ Method

HiWISE™ is modeled on Latent Semantic Analysis (LSA, Landauer and Dumais, 1997) and Hyperspace Analogue to Language (HAL, Burgess and Lund, 1997). All three algorithms utilize word co-occurrence statistics to yield a large, high dimensional matrix which characterizes general language usage. In this matrix, each row (vector) represents a particular English word from a given dictionary. The raw data is transformed so that these vectors can be compared mathematically, on a like for like basis, using a dot product or Euclidean distance measure. These techniques can be validated by correlation to psychological measures of semantics, for example judgments of semantic similarity between word pairs. LSA and HAL both organize information is such a way that they can achieve impressive correspondence to humans on this and a range of other tasks involving lexical semantic knowledge.

HiWISE™ creates a lexical co-occurrence matrix from a plain text corpus. It works by scanning a corpus word by word for instances of a set of target words. It scores the target instance against any of a smaller set of '*hotwords*' that occur in the immediate vicinity. The size of this vicinity (*context window*) is **w** words before and after each instance in the corpus. Each co-occurrence (hit) found in the context window increments the matrix element(s) corresponding to the particular target and vector word combination.

In HiWISE™, the target word set is usually all of the English words in a dictionary (e.g. 60,000 words). The hotword set is an independent set of **n** (<2000) English words that yield the most informative characterization of targets. HiWISE™ generates two matrixes, termed the '*tall*' and '*long*' matrix. Each row in the tall **m**x**n** matrix output file represents co-occurrences of hotwords **before** a particular target in the corpus, while each column in the long **n**x**m** matrix represents co-occurrences **after** it.

The program works by looking up each word of the corpus in both the target list and the hotword list. HiWISE™ relates the current word to each cell in the context window. The context window cells contain both a target number and a vector set number for the words that precede the current word. If the current word is a target, and the window word is a vector word, then the cell in the tall matrix is incremented. If the reverse is true (current word is a vector word, window word is a target), then the relevant cell in the long matrix is incremented.

At the end of corpus analysis, the long matrix is transposed and concatenated to the tall matrix to generate a single **m** x 2**n** dimension matrix. Several mathematical techniques are applied to the matrix. First the *expected values* for each cell are subtracted; this is the frequency at which the target word and hotword could be expected to co-occur, given their independent frequency. If the resulting value is positive, then they are associated to a greater degree than expected, conversely, if the result is negative they are dissociated. The sum of these values (squared) for a given hotword (column) indicates how distinctive (and hence informative) it is. This method is used to select the most informative hot words.

Finally, the vectors are normalized – this adjusts for the differing frequencies of particular words, and ensures that vectors are combined and compared on a unitary basis.

EVALUATION

Evaluation of the HiWISE Algorithm

IAI developed a semantic analysis technology (using IR&D funds) called HiWISE™, and evaluated it for potential use in the summary assessment module of GradAtions. HiWISE™ (Word Information Semantic Engine), generates a statistical representation of each word in a dictionary, based upon the contextual patterns that they occur in a natural language corpus. Via an agreement with MetaMetrics, Inc., we used their corpus of 500 million words to train the algorithm. Both HiWISE™ and LSA (Landauer & Dumais, 1997) were then compared to a benchmark word matching measure ('Dupes'), which is simply the proportion of summary words present in the original passage, and to trained human scoring. Semantic analysis algorithms

offer two key advantages over word matching. First, they allow for the use of novel words in the summary that have similar meanings to words in the passage. Second, they score summaries according to their relevancy to the whole passage– allowing us to measure how well the shorter text characterizes the most important aspects of the longer text.

These advantages allow semantic analysis methods to achieve a more accurate assessment of summary quality. To evaluate these hypotheses, we conducted a study to measured HiWISE™ correspondence to trained human graders, and compare these scores to LSA and Dupes. Ten human graders were asked to score 58 summaries on a 1-5 scale, using a predefined scoring rubric. These scores were then correlated to the artificial scores. Overall HiWISE™ correlation to the mean human grade was highly significant ($r=0.47$, $p<0.0005$), superior to both LSA ($r=0.22$, $p<0.05$) and Dupes ($r=0.215$, $p<0.05$).

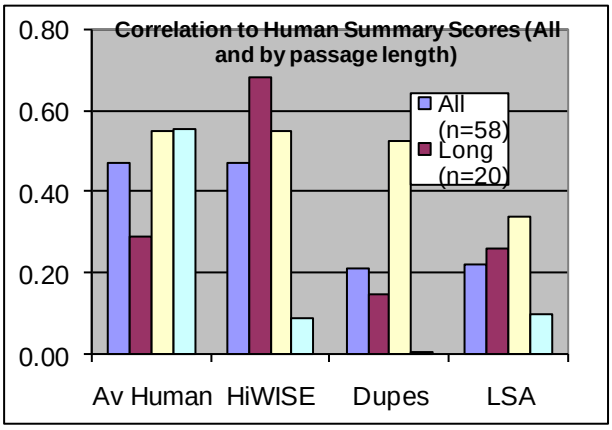

Figure 2.

However, these figures mask discrepancies in scoring of passages of different lengths. To examine these effects, the passage/summary pairs were divided into three groups (long, medium and short summary length). Humans were most consistent with the 18 shorter summaries ($r=0.55$, $p<0.01$), whereas HiWISE™ failed to achieve correspondence ($r=0.09$, NS). For the 20 longest summaries (greater than 200 words), HiWISE™ achieves a high correlation ($r=0.68$, $p<0.0005$), while humans, LSA and Dupes fair worse ($r=0.29$,$r=0.26$,$r=0.15$ respectively, all NS). Correlations for medium length passages were very similar for humans, HiWISE™ and Dupes ($r=0.55$,$r=0.55$, $r=0.52$ respectively, all $p<0.01$), while LSA score $r=0.34$. These results are illustrated in the bar chart above.

The results indicate that HiWISE™ achieves improved accuracy over LSA for medium and longer passages, and improved accuracy over Dupes for longer passages. None of the computational methods provided accurate grading for shorter summaries.

Evaluation of Learning Gains

Preliminary study. The first study concluded by us was a usability study that showed that 93% of the participants were able to successfully navigate through the GradAtions™ system and, although this study did not empirically measure learning gains from using the system, 86% of the participants reported that they learned new reading strategies that would help them in the future. 81% reported that they enjoyed using GradAtions.

Field study 1: IAI conducted a field study at the Navy Great Lakes Training Center using a pre-test/post-test design with 160 subjects who were enrolled in a required academic skills remediate program prior to attending one of several Navy A Schools. The research protocol required subject to use GradAtions for 35 hours over a period of 4-6 weeks; however data logs from the system revealed that the mean hours of use were 17, with too wide a range to make pre-test/post test analysis meaningful. However, in a post-test only measure, students who used GradAtions scored significantly higher on measures of identifying key words and recognizing main ideas in Navy texts than students in a control group receiving classroom instruction only.

Field Study 2. Gradations was evaluated by Dr. Robert Atkinson, in an evaluation funded by ONR at the Great Lakes training facility. In that independent evaluation 3 computers based reading tutors were evaluated. (*Read-On!, STAR,* and *Gradations*) (Atkinson, 2007). The evaluation was aimed to explore learning gains associated with each of the tutors, compare learning gains across each pair of the tutors, and explore learning gains associated with using all three tutors, or alternatively determines the optimal ordering of the tutors. Learning profile and average amount of time learner spent on the tutor was also to be evaluated.

Students were trained on various combinations of the three tutors and in different order. Students were also compared against the current commercial reading comprehension currently being utilized by the U.S. Army (*Lifetime Learning's Reading Program*). The experiments indicated that participants that initially worked exclusively on Gradations and Read-On! produced significantly higher scores after working with

these tutors. On the other hand, STAR and Lifetime Learning did not produce significant learning gains after the initial presentation of the software.

As for comparing sequences of using the tutors, there were four combinations of the three reading programs that produced significantly higher scores, of these, the Gradations → STAR→ Read-On! combination produced the largest learning gains. The participants that worked with the Gradations → STAR → Read-On! sequence had a gain of 15.5 points, which corresponds with a reading comprehension grade level gain of 1.1. Thus, the most optimal way of ordering the three tutors is Gradations first, STAR second, and Read-On! third.

Assessment 4: Adjusted (for location) means (M), standard errors (SE), reading grade level equivalents (GL) for ACCUPLACER 1 and 4, gain scores (means and grade level equivalents) across conditions for all participants (n =159).

	Accuplacer 1			Accuplacer 4			Gain	
Condition	M	SE	GL	M	SE	GL	M	GL
Read-On! → STAR→ Gradations	50.2	4.4	8.7	56.7	5.2	9.0	+6.5	+0.3
Read-On! → Gradations → STAR	48.2	6.6	8.7	53.6	7.8	9.0	+5.4	+0.3
STAR → Read-On! → Gradations	49.4	6.2	8.7	51.7	7.4	8.8	+2.3	+0.1
STAR → Gradations → Read-On!	46.8	4.1	8.5	55.3	4.9	9.0	+8.5	+0.5
Gradations → Read-On! → STAR	43.2	4.6	8.2	52.5	5.5	9.0	+9.3	+0.8
Gradations → STAR → Read-On!	48.2	5.2	8.7	63.7	6.1	9.8	+15.5	+1.1
Lifetime Learning	56.3	6.3	9.0	54.5	7.5	9.0	-1.8	0

Figure 3. Best learning gain was achieved when students used GradAtions first.

Field Study 3**:** A third field study is currently under way at the Navy Great Lakes Recruit Training Center. To date, approximately 100 recruits have participated in the study that explores the effectiveness of instructional interventions designed to improve recruits' reading comprehension ability. Preliminary results show that students using GradAtions and two other tutors had an average gain 2.06 grade levels, which was significantly different from students using a classroom-based approach.

DISCUSSION

The success of GradAtions is derived from the intersection of technical capability (automated text analysis in this case) and theoretical research (summarizing and reading comprehension). The evaluations done with GradAtions, show that it is highly effective at improving adult reading capabilities. Its primary means of assessing student learning by having students write summaries. Future work with GradAtions could expand its capabilities by allowing students to demonstrate mastery of materials in other ways as well. For example, using speech to text technology, students could provide oral summaries as well as written summaries that could then be scored by GradAtions. Another direction is to allow content authors to create questions for students to answer as is typical in many standardized testing settings.

Perhaps one of the most intriguing extensions of GradAtions would result from customizing GradAtions to specific domains. Currently, GradAtions is designed as a generic reading tutor that accommodates any type of reading materials. As such, its assessment of student comprehension is limited to the form of the summaries produced by students rather than in-depth analysis of their content.

Within the Navy and other military organizations, there can be two goals associated with building reading skills in students. The first is to boost specific literacy skills and the current version of GradAtions is well suited for that. The second is to boost proficiency in the performance of specific duties. Often military organizations use simulators to train procedural tasks, but there could still be factual or conceptual components to these tasks for which text is the preferred medium of communication.

In such cases, a blended learning environment of text-based and simulation-based instruction may provide the optimal solution. To optimize instructional effectiveness, not only should the text-based component of the instruction support the simulation-based component, performance on the latter may be a more useful measure of how well a student learned the material presented in the former.

Therefore, one can imagine text-based instruction authored in and delivered by GradAtions followed by a simulation-based scenario in which the student has the opportunity to demonstrate his or her proficiency in the learned subject matter. Here, instead of a summary rating being used by GradAtions to determine mastery of the subject matter and subsequent instruction, performance on the simulated-scenario would be recorded and provided to GradAtions as the basis for determining future instruction. If implemented in this way, GradAtions can be applied more directly to specific end user jobs and the assessments would be more diagnostic of actual readiness to perform a job as a result of training.

We wish to acknowledge the helpful guidance and advice provided by Dr. Susan Chipman, formerly of the Office of Naval Research for her insights and contributions to this work.

REFERENCES

Aarnoutse, C., Brand-Gruwel, S., Oduber, R. (1997), Improving Reading Comprehension Strategies through Listening, *Educational Studies*, *v23 n2, p209-27.*

Armbruster, B.B., & Anderson, T.H. (1984), The effects of mapping on the free recall of expository text, *Technical Report No. 160, University of Illinois, Center for the Study of Reading.*

Armbruster, B.B., Anderson, T.H. & Osertag, J. (1987), Does text structure/summarization instruction facilitate learning from expository text?, *Reading Research Quarterly*, v22, p 332-346.

Arnold, H. F., (1942) The comparative effectiveness of certain study techniques in the field of history, *Journal of Educational Psychology*, *v33, p 449-457.*

Atkinson, R.K. (2007). An Experimental Evaluation of Three Computer-Based Reading Comprehension Tutors. *Final Report to Office of Naval Research.*

Bean, T. W., & Steenwyk, F.L. (1984), The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension, J*ournal of Reading Behavior, v15, p297-306.*

Brown, A.L., & Day, J.D. (1983), Macrorules for summarizing texts: The development of expertise, *Journal of Verbal Learning and Verbal Behavior*, v22*, n1, p1-14.*

Burgess, C., & Lund, K., (1997), Modeling Parsing Constraints in High-dimensional Context Space, *Language and Cognitive Processes*, *vol. 12 (2/3), pp. 177-210.*

Carter, Carolyn J. (1997), Why Reciprocal Teaching?*, Educational Leadership, v54 n6, p64-68.*

Duke, N.K., & Pearson, P.D. (1998). Effective Practices for Developing Reading Comprehension,. *The Rand Corporation.*

Germane, C.E. (1921) Outlining and summarizing compared with reading as methods of studying. In G. M. Whipple (Ed.), *20th Yearbook of the National Society for the Study of Education*, *Part II. Bloomington, IL: Public School Publishing Co.*

Gilroy, A. Moore, D.W. (1998), Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities with Ten Primary School Girls, *Educational Psychology*, *v8 n1-2, p41-48.*

Gude, S., Jackson, T., & Shaw, A. (2000), Strategies for Improving Reading Comprehension in Content Areas*, Master of Arts Action Research Project*, Saint Xavier University, Available via EDRS Service (ref ED441229).

Hare, V.C., Borchardt, K.M. (1984) Direct instruction of summarization skills, *Reading Research Quarterly,* v20, p 62-78.

Head, M.H., Buss, R.R. (1987) Factors affecting summary writing and their impact on reading comprehension assessment, In J.E. Readence & R. S. Baldwin (Eds.), *Research in literacy: Merging perspectives* (36th Yearbook of the National Reading Conference) *p25-33.* Rochester, NY: National Reading Conference.

Landauer, T.K., & Dumais, S.T., (1997), A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge', *Psychological Review*, vol. 104 (2), pp. 211-240

Loranger, A.L. (1997), Comprehension Strategies Instruction: Does It Make a Difference?, *Reading Psychology, v18 n1, p31-68.*

Rinehart, S. D., Stahl, S.A., Erickson, L.G. (1986) , Some effects of summarization training on reading and studying, *Reading Research Quarterly,* v 21, p422-438.

Taylor, N.E., Wade, M.R., Yekovich, F. R. (1985), The effects of text manipulation and multiple reading strategies on the reading performance of good and poor readers, *Reading Research Quarterly,* v 20, p 566-576.

van Dijk, T.A., & Kintsch, W. (1978) Cognitive psychology and discourse: Recalling and summarizing stories in W. U. Dressler (Ed.) *Current trends in textlinguistics* (p 127-147). New York: de Gruyter.