Psychology Faculty Publications

Psychology

2002

Using Latent Semantic Analysis to Assess Reader Strategies

Joseph P. Magliano

Katja Wiemer-Hastings

Keith K. Millis

Brenton D.. Muñoz

Danielle NcNamara
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_fac_pubs

Part of the Experimental Analysis of Behavior Commons, and the Quantitative Psychology Commons

Using latent semantic analysis to assess reader strategies

JOSEPH P. MAGLIANO, KATJA WIEMER-HASTINGS, KEITH K. MILLIS, and BRENTON D. MUÑOZ
Northern Illinois University, DeKalb, Illinois

and

DANIELLE MCNAMARA
Old Dominion University, Norfolk, Virginia

We tested a computer-based procedure for assessing reader strategies that was based on verbal protocols that utilized latent semantic analysis (LSA). Students were given self-explanation–reading training (SERT), which teaches strategies that facilitate self-explanation during reading, such as elaboration based on world knowledge and bridging between text sentences. During a computerized version of SERT practice, students read texts and typed self-explanations into a computer after each sentence. The use of SERT strategies during this practice was assessed by determining the extent to which students used the information in the current sentence versus the prior text or world knowledge in their self-explanations. This assessment was made on the basis of human judgments and LSA. Both human judgments and LSA were remarkably similar and indicated that students who were not complying with SERT tended to paraphrase the text sentences, whereas students who were compliant with SERT tended to explain the sentences in terms of what they knew about the world and of information provided in the prior text context. The similarity between human judgments and LSA indicates that LSA will be useful in accounting for reading strategies in a Web-based version of SERT.

Successful students engage in specific strategies when reading difficult texts, such as explaining, using logic, and elaborating. Because many students do not use these strategies and are poor readers, McNamara and her colleagues developed Self-Explanation Reading Training (SERT), which teaches active reading strategies (McNamara & Scott, 1999). SERT emphasizes several strategies that improve the process of self-explanation. Self-explanation refers to the act of explaining difficult text to oneself. The strategies include using logic or world knowledge to elaborate the current sentence, making conceptual bridges among ideas in the text, and predicting what will come next in the text. Not only has SERT been shown to promote general reading comprehension, it has been shown to improve overall class performance, particularly for poor students (McNamara & Scott, 1999).

We are developing a Web-based version of SERT that will enable many students to take advantage of this training. Of course, implementing SERT poses various challenges, one of which is to make a computer "understand"

students' self-explanations. This is crucial to the Web-based SERT, because a central component of SERT is a practice session, in which students work in pairs reading a difficult text, encouraging one another to use the SERT strategies in forming their self-explanations. In the Web-based version of this component, students will read difficult scientific texts and type in their self-explanations after each sentence. When needed, an animated agent will supply feedback on the quality of the self-explanations. Feedback concerning the quality of self-explanation during training will be guided by latent semantic analysis (LSA; Landauer & Dumais, 1997).

The goal of the present study was to test one approach to using LSA to assess the extent to which students are using the strategies emphasized by SERT in the Web-based trainer. We will refer to a student who shows multiple strategies in their self-explanation as *complying* with SERT; a student who merely paraphrases the current sentence or types in something vague will be said to be *noncompliant* with SERT. We administered SERT in a traditional classroom setting, which included a practice session that occurred in the classroom. After SERT was administered, the students were invited to engage in additional practice of the SERT strategies on a computer. During this additional practice, they typed in self-explanations after reading each sentence of two texts. We compared an assessment of compliance with SERT based on human judgments with one based on LSA.

Before we describe the LSA-based approach that we are exploring, it is important to first explain LSA. LSA is a text-

processing tool that represents the semantic contents of text units on the basis of their co-occurrence frequency with all other text units within a large corpus of text. First, LSA computes a matrix of how frequently individual words co-occur with each other within all documents of text in the database. The matrix is then transformed; an algorithm called singular value decomposition is applied to the matrix to reduce the dimensionality to an "ideal" number. This number is determined empirically by assessing how well LSA text evaluations match the evaluations of domain experts. In the resulting high-dimensional semantic space, each text unit is represented as a vector with as many elements as there are dimensions. When presented with two text units, LSA computes their similarity by computing the cosines of their vectors. The cosine measures the similarity of the two vectors across all dimensions. The more similar the vectors are, the higher the LSA cosine is. Cosines of 1 indicate maximal similarity. The minimal cosine is 0 and indicates maximal dissimilarity. LSA cosines of text units, both words and paragraphs, have been shown to reliably match human similarity judgments of documents (Landauer & Dumais, 1997; Landauer, Laham, Rehder, & Schreiner, 1997).

Self-Explanation Reading Training and Sample Self-Explanations

SERT was inspired by previous research showing the benefits of strategy instruction (Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994; Magliano, Trabasso, & Graesser, 1999; Palinscar & Brown, 1984; Yuill & Oakhill, 1988). Training begins with a brief instruction that includes definitions and examples of reading strategies associated with self-explanation (see Protocol 1 in Table 1 for an example of a self explanation). The strategies involve making bridging inferences between separate ideas in the text, using prior knowledge and logic to understand the text, predicting what the text will say, and monitoring comprehension. After this brief instruction, students read a science text and watch a video of a student in the process of self-explaining the text. At certain points in the text, the students identify the strategies used by the student in the video and then discuss these strategies as a

group. In a final stage of SERT, the students work in pairs to practice strategies by taking turns reading out loud text sentences and sharing self-explanations. Instructors are present to assist and monitor the students.

In order to get a sense of how self-explanations can reflect differential compliance with SERT, consider Table 1, which contains sample self-explanations produced while reading a text titled "Heart Disease" (see the Appendix for the entire text). These self-explanations were generated to the sentence "It (blood) becomes purplish, and the baby's skin looks blue." Self-explanation 1 reveals a number of strategies advocated in SERT: elaborations based on world knowledge (e.g., the statements pertaining to "choking") and bridges to prior text information (e.g., "not receiving enough oxygen" and "heart disease"). In contrast, the reader who generated Self-explanation 2 bridged the sentence to the immediately prior sentences regarding "carbon dioxide" but did not discuss how the current sentence was related to the general topic of heart disease or provide an elaboration based on relevant world knowledge. Finally, the reader who generated Self-explanation 3 merely "parroted" the current sentence by paraphrasing it, which would not be considered as complying with SERT.

These examples reflect different types of reading strategies proposed by Coté and Goldman (1999). We used their typology of reading strategies to assess compliance with SERT. A *knowledge-building* explanation includes how the sentence is related to the student's world knowledge, the prior text, and to the general message or theme of the text. In giving this type of explanation, a student tends to utilize multiple reading strategies emphasized in SERT. Self-explanation 1 is an example of knowledge building. A *sentence-focused* explanation focuses primarily on the sentence. The student might elaborate upon a concept in the sentence or talk about how the sentence is related to the immediate prior sentence but does not explain how the sentence is related to the overall message of the text. Self-explanation 2 is an example of a sentence-focused explanation. Sentence-focused explanations reflect only partial compliance with SERT, because SERT emphasizes self-explanations that link a sentence to the overall message of a text. Finally, a *minimalist* explanation is one that para-

Table 1
Example Knowledge-Building, Sentence-Focused, and Minimal SERT Self-Explanations for the Sentence "It (Blood) Becomes Purplish, and the Baby's Skin Looks Blue," From the Text "Heart Disease"

Protocol	Reading Strategy	Clause	Source
1	Knowledge-building	1. This gives the impression of someone choking	WK
		2. When someone chokes,	WK
		3. they start to turn colors,	WK
		4. and the infant is essentially choking from the inside.	WK
		5. The skin turning blue	CS
		6. might have something to do with not receiving enough oxygen,	PT
		7. connected to the heart problems.	PT
2	Sentence-focused	1. When the carbon dioxide does not escape the body	PT
		2. the baby's skin looks blue.	CS
3	Minimal	1. The blood turns to a purplish color,	CS
		2. and the baby's skin turns blue.	CS

Note—WK, world knowledge; CS, current sentence; PT, prior text.

phrases the sentence or is vague (e.g., OK). Self-explanation 3 is an example of a minimal explanation.

Using LSA to Assess Compliance with SERT

We tested whether LSA can classify self-explanations as knowledge building, sentence focused, or minimalist. The present approach involved calculating a measure of semantic similarity between a student's self-explanation and *semantic benchmarks* associated with that sentence. LSA provides the measure of semantic similarity. Semantic benchmarks represent information from different sources that a reader could be drawing upon in producing the protocol. In this context, they are merely a collection of words. The semantic benchmarks refer to the (1) current sentence, (2) causally important prior sentences, and (3) relevant world knowledge and represent the different sources that the reader can draw upon when self-explaining.

This type of approach has been used successfully in predicting comprehension differences between skilled and less skilled readers (Magliano & Millis, 2000). Furthermore, Graesser and his colleagues have successfully used LSA in a computerized tutor called AutoTutor (Graesser et al., 2000; Wiemer-Hastings, Wiemer-Hastings, & Graesser, 1999). With AutoTutor, students are asked questions by a computerized tutor and are required to type their answers into the computer. AutoTutor determines the degree of correctness of the answers by using LSA to determine the semantic overlap between them and the ideal answers. AutoTutor provides feedback to a student on the basis of the magnitude of the cosine values produced by this analysis. Our use of LSA is conceptually similar to that of Graesser and his colleagues.

Knowledge-building self-explanations should have a high overlap (i.e., high LSA cosines) with causally important information from the prior text and/or relevant world knowledge. In contrast, a minimalist self-explanation should have a relatively low overlap with the prior text and relevant world knowledge, but a relatively high overlap with the current sentence, because the reader is primarily paraphrasing the current sentence. A sentence-focused self-explanation should also have a relative high overlap with the current sentence but should have intermediate overlap with the prior text and relevant world knowledge.

We adopted a two-step procedure to assess whether LSA could be reliably used to assess compliance with SERT in this manner. First, we conducted an assessment based on human judgments. In this step, raters determined the number of clauses in a self-explanation that were based on the current sentence, prior text information, or general world knowledge. These constituted the *sources* of the information mentioned in an explanation. Raters also classified the self-explanations on the basis of whether they depicted knowledge-building, sentence-focused, or minimalist responses. These constituted *reading strategies* conveyed by an explanation. We then verified the assumption that different reading strategies drew upon different sources of information. In the second step, we used LSA as a surrogate for human raters, in the manner described in the preceding paragraph. Finally, we predicted reading strategies from the LSA cosines in order to determine whether this approach could identify reading strategies and whether a user of the Web-based SERT would be complying with SERT.

METHOD

Participants

Two hundred and twelve undergraduates from Northern Illinois University participated for course credit. Forty participants were enrolled in a critical thinking course. These students received SERT as part of the critical thinking course. One hundred and seventy-two of the participants were enrolled in an introduction to psychology course. These students did not receive SERT. However, these students provided verbal protocols for the construction of the semantic benchmarks used in the LSA analysis.

Procedure

SERT was administered across 2 consecutive days to an undergraduate critical thinking course ($n = 40$). The administration of SERT followed a script that was developed by McNamara (McNamara & Scott, 1999). This script consists of three training modules. The first module was strategy introduction, which lasted approximately 25 min. During strategy introduction, the participants were given definitions and examples of the strategies associated with self-explanation: comprehension monitoring, paraphrasing, elaboration, logic/reasoning, bridging, and prediction. The examples consisted of sentences taken from scientific texts and self-explanations produced with those sentences. The second module involved a modeling of SERT practice, in which the participants viewed a videotape of a student practicing the SERT strategies. The student read a text out loud, one sentence at a time, and practiced the SERT strategies by thinking aloud. The students in the course followed along with a written transcript of the videotape. The videotape was stopped at six preselected sentences, at which time the instructor invited the students to discuss the strategies that were demonstrated. The third module was practice, which took place during the second class period. During practice, the students were grouped into pairs. They were given a practice text and were instructed to take turns self-explaining each sentence in the text. Rather than thinking aloud, as in the videotape, the students wrote their self-explanations on sheets of paper. After each sentence, the students were instructed to identify and evaluate the use of the SERT strategies that were used by the student who was practicing with that sentence.

Within a week of SERT training, these participants were tested individually, providing self-explanations for scientific texts that were presented on a computer. These self-explanations served as the primary data analyzed in this study. The participants were instructed to type in a self-explanation after reading each sentence of two texts. The texts were presented sentence by sentence in one box on the screen, and the participants typed their self-explanations into another box. When they first clicked a "next" button, the title of the text appeared in the text box. For the title, the students were instructed to type in a prediction for the first sentence. After the student typed their predictions in the response box, they clicked the "next" button again, and the first sentence of the text appeared. They then typed in their self-explanations to the first sentence, and the next sentence immediately was added to the text after they clicked the "next" button. They then typed in their self-explanations for that sentence. The students progressed in this fashion until they had read two texts. The computer recorded all responses. Paragraph formatting was maintained in the presentation of the text so that the text would look natural to the participants. The participants could use the scroll bar to reread any portion of the text that was not visible on the screen. One half of the participants read the texts on the development of coal and

heart disease, whereas the other half read texts on the development of thunderstorms and the food pyramid. The passages were moderately difficult to read, suitable for freshman college students, and ranged between 20 and 34 sentences long (total $n = 98$). An equal number of participants read each passage, and the order of the passages was counterbalanced across participants.

Coding self-explanations on reading strategy. We chose a sample of 36 sentences to analyze (i.e., 37% of the sentences across the four stories). Sentences were included in the sample if (1) at least 25% of the self-explanations were classified as knowledge building (see below) and (2) there were semantic benchmarks for both the prior text and world knowledge (see below). Two independent raters categorized the self-explanations for these sentences as a minimal explanation, a sentence-focused explanation, or a knowledge-building explanation. In order to make this decision, the raters first parsed the self-explanations into clauses containing a main verb (Table 1 presents example clauses). Minimal explanations contained only causes that either were vague or were partial or entire paraphrases of the current sentence. Sentence-focused explanations usually contained paraphrases as well but included at least one clause that contained either an elaboration based on world knowledge or a *bridge* from the current sentence to the prior sentence. Knowledge-building explanations contained multiple clauses that were elaborations from world knowledge or bridges from the current sentence to prior text sentences or to the theme of a text (e.g., heart disease). As such, knowledge-building explanations reflected the use of multiple SERT strategies, in addition to paraphrasing. There were 291 knowledge-building, 235 sentence-focused, and 160 minimalist self-explanations across the 36 target sentences. Interrater reliability in determining reading strategies was high (kappa = .91)

Coding self-explanations on informational source. The explanations were also categorized in terms of what informational sources contributed to their content. Three sources were considered: the current sentence, prior text, and world knowledge. Clauses based on the current sentence generally restated or paraphrased a clause in the current sentence (e.g., Clauses 1 and 2 of Self-explanation 3). Clauses based on the prior text reinstated or paraphrased a sentence or concept that was explicitly stated in the prior text (e.g., Clause 1 from Protocol 2 and Clauses 6 and 7 from Self-explanation 1). Clauses based on world knowledge contained information not explicitly mentioned in the current sentence or prior text sentences and were assumed to come from the world knowledge of the student (e.g., Clauses 1–4 in Self-Explanation 1). Interrater reliabilities in judging the sources of the clauses were high (alpha = .92, .93, and .92 for current sentence, prior text, and world knowledge, respectively).

Constructing semantic benchmarks. As was mentioned earlier, the semantic benchmarks were groups of words that we compared with the self-explanations via LSA in order for the computer to assess reading strategy. Three benchmarks were constructed for each sentence of the four experimental texts: current sentence, prior text, and world knowledge. The current sentence benchmarks consisted of content words[1] in the sentence (i.e., nouns, main verbs, adjectives, and adverbs). The prior text benchmark contained words from the prior text that were important, either theoretically or empirically, to the current sentence. Theoretically important words were identified via a causal network analysis (CNA; Trabasso, van den Broek, & Suh, 1989). CNA determines causal relationships among sentences (see Trabasso et al., 1989, for a detailed discussion on the

criteria for conducting a CNA). CNAs were conducted on the texts by the first author. The theoretically important words were taken from previous sentences that were directly causally connected to the sentence and were not in the current sentence benchmark. The empirically important words were additional content words related to the prior text. These were gleaned from verbal protocols produced by a separate group of participants who did not receive SERT (see below). To be included in the benchmark, a word must have been produced by 2 or more of these participants. Finally, the world knowledge benchmarks consisted of words produced by 2 or more these additional participants to the current sentence. They were content words that were (1) produced for a sentence that were (2) not in the current sentence or in a prior text sentence and not close synonyms of the words in the sentences. Table 2 contains the benchmarks for the sentence "It (blood) becomes purplish, and the baby's skin looks blue," from the text "Heart Disease."

As was mentioned above, there was an independent group of participants ($n = 172$) who produced protocols to the texts. These participants were given one of four instructions. The instructions emphasized different strategies related to SERT and, together, were thought to elicit a maximal amount of world knowledge associated with the text sentences. One fourth of the participants were told to use their general knowledge of the world to elaborate each sentence in the texts. One fourth of the participants were told to explain the text sentences on the basis of information provided in prior text sentences. Another fourth were instructed to predict or anticipate what the author would discuss next. The last fourth were told to restate the sentences in their own words. All of these participants were given practice texts and feedback on their practice responses. For each sentence, we collapsed all of the responses across the four instructional groups.

Obtaining LSA cosines. The University of Colorado LSA Web site was used for the LSA analysis (HYPERLINK http://lsa.colorado.edu/) http://lsa.colorado.edu/). The Colorado Web site contains different document spaces based on topic (e.g., general reading, psychology, heart, etc.). The topic space that was used in the present study was general-reading-up-to-the-first-year-in-college , with 300 factors. This space contains a large sample of texts that first year college students should have been exposed to before entering college. We chose this space because we believed that it would best reflect the general knowledge needed to understand our practice texts, as well as the general knowledge of the readers in our participant population. For every sentence of our sample, we obtained the LSA cosine between every self-explanation that was produced and the three benchmarks. It is important to note that the entire self-explanation was submitted, not individual clauses. The comparison type was document-to-document, which is appropriate when comparing text units larger than individual words. Because our goal was to assess whether LSA could distinguish between SERT-compliant and SERT-noncompliant explanations, we summed the cosines for prior text and world knowledge, both of which would be considered as complying with SERT.

RESULTS

Human Ratings

For each of the 37 sentences, we calculated the proportion of clauses falling into each source and each reading

Table 2
Example Current Sentence (CS), Prior Text (PT), and
World Knowledge (WK) Benchmarks

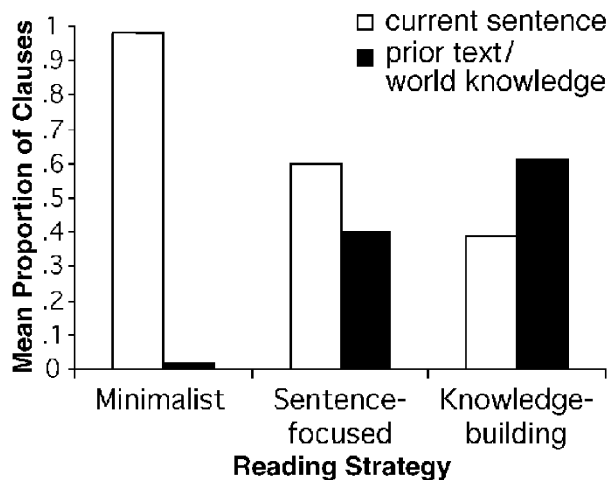| Text Sentence	CS	PT	WK
It (blood) becomes purplish, and the baby's skin looks blue	blood purple baby skin blue	rid carbon dioxide lungs receive oxygen heart body	turns color lack excess need result amount attention die

Figure 1. The mean proportion of idea units based on current sentences and prior text/world knowledge for minimal, sentence-focused, and knowledge-building self-explanations.

strategy. Therefore, the sentence, not participants, served as the unit of analysis. The proportions for prior text and world knowledge were summed so that they would be consistent with the LSA analysis. Figure 1 shows the proportion of clauses that were from the current sentence and the proportion of clauses that were either from the prior text or from world knowledge for each type of strategy. The means were submitted to a 3 reading strategies (knowledge building, text focused, minimal explanation) $\times$ 2 sources (current sentence vs. prior text or world knowledge) within-sentence analysis of variance (ANOVA). As was predicted, there was a significant interaction between reading strategy and source [$F(1,35) = 771.10$, $MS_e = 0.01712$, $p < .01$]. The proportion of clauses that contained information from the current sentence decreased significantly from minimalist to sentence-focused explanations and from sentence-focused to knowledge-building explanations. In contrast, the proportion of clauses that came from prior text or from world knowledge increased significantly from minimalist to sentence-focused explanations and from sentence-focused to knowledge-building explanations.[2]

The importance of these findings is that they confirm the assumption that different types of reading strategies rely on different informational sources. Reading strategies emphasized by SERT rely on information from world knowledge, prior text, and the current sentence. A passive reading strategy, on the other hand, merely requires access to the current sentence.

LSA Cosines

For each of the 36 sentences, we computed the mean LSA cosines for each type of self-explanation (reading strategy) and source. Figure 2 shows the resulting means. The means showed the same pattern as the hand-coded data presented in Figure 1. The ANOVA revealed a significant strategy $\times$ source interaction [$F(1,35) = 78.57$, $MS_e =$

0.008398, $p < .01$]. The pattern of significant differences reported for the proportion data was found here. These findings are important because they correspond to the hand-coded judgments, further indicating that LSA can be used to reliably code text. They are also important because they indicate that LSA can be used to detect the source of information in self-explanations, which will be critical for the Web-based SERT's to identifying whether the student is typing in reasonably good self-explanations.

Predicting Reading Strategies

We used discriminant analysis to predict the reading strategies for the self-explanations (minimalist, sentence focused, knowledge building) from the LSA cosines between the explanations and the benchmarks for the current sentence and from the sum of the cosines for prior text and world knowledge. Two discriminate functions were calculated, although the first accounted for 99% of the between-strategy variability [$\chi^2(6) = 98.8$, $p < .001$]. The functions were able to correctly classify 47% of the self-explanations: 67%, 18%, and 58% of the minimalist, sentence-focused, and knowledge-building explanations, respectively. As one can see, with LSA values, we had the most difficulty accounting for sentence-focused explanations. LSA was able to do significantly better than chance (i.e., 33%) for minimalist and knowledge-building explanations. We also added mean vector length in the self-explanation as a predictor. Vector length is an indicator of how much information LSA has about a word or, in this case, the entire set of words in a self-explanation (Kintsch, 2001). Because vector length is correlated with the number of words in the entry, its inclusion in the model can be interpreted to partial out the effect of self-explanation length. The mean vector lengths for minimalist, sentence-focused, and knowledge-building explanations were 6.1, 7.6, and 13.0, respectively (the corresponding mean numbers of words were 12, 16, and 34). When we added vector length as a
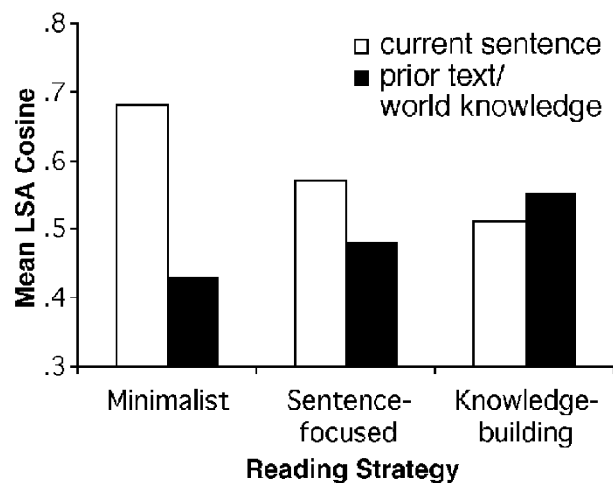

Figure 2. The mean cosine values for current sentences and prior text/world knowledge for minimal, sentence-focused, and knowledge-building self-explanations.

Table 3
Logistic Regression Coefficients (*B*) and Standard Error (*SEs*) for Predicting Compliance With SERT
From LSA-Based Variables, With and Without Words

	Minimalist/Sentence-Focused Versus Knowledge-Building		Minimalist Versus Knowledge-Building	
	B	*SE*	*B*	*SE*
LSA-based				
Current sentence	−1.65***	−2.78***	−3.21***	−3.70***
Past sentence/world knowledge	2.05***	1.16**	4.17***	2.13**
Mean vector length in explanation	not included	.43***	not included	.35***
Nagelkerke R^2	.11***	.53***	.28***	.69***

Note—Complying with SERT was coded as 2; noncompliance was coded as a 1.   **$p < .05$.   ***$p < .001$.

predictor variable, two significant functions were calculated, accounting for 96% and 4% of the between-strategy variance ($ps < .01$). This equation was able to correctly classify 60% of all self-explanations: 69%, 45%, and 67% of the minimalist, sentence-focused, and knowledge-building explanations, respectively. Vector length dramatically improved classification.

Predicting Compliance With SERT

The bottom-line test of our approach is whether LSA cosines predict whether the user is typing in self-explanations that employ the SERT strategies. Therefore, we coded minimalist and sentence-focused self-explanations as being noncompliant with SERT and knowledge-building self-explanations as complying with SERT. We then used logistic regression to predict compliance from the cosines between the self-explanations and the semantic benchmarks. The resulting equation was significant [$\chi^2(2) = 57.24$, $p < .001$]. Each of the predictor variables was significant ($ps < .01$; see the left-hand side of Table 2 for the regression coefficients). The equation correctly classified 61% of the explanations: 80% and 36% of the noncompliant and compliant explanations, respectively. When vector length was included in the equation, 79% of the explanations were correctly classified: 87% and 72% of the noncompliant and compliant explanations, respectively. For this equation, all the predictors were statistically significant. Again, vector length proved to be a robust predictor of whether a self-explanation contained multiple strategies (i.e., knowledge-building). Vector length increased the percentage of correctly classified noncompliant explanations 7% above and beyond the LSA-based predictors but increased it an impressive 36% for compliant explanations.[3]

We also conducted an analysis that predicted minimalist versus knowledge-building explanations. This analysis excluded sentence-focused explanations. This was warranted because sentence-focused explanations were slightly ambiguous as to whether they reflected compliance with SERT. On the one hand, they contained only one reference to a prior sentence or world knowledge, indicating that the student was not engaging in multiple SERT strategies. But on the other hand, there was at least some strategy use beyond merely paraphrasing the sentence. In this sense, they were using SERT. The logistic regression predicting minimalist versus knowledge-building explanations from the LSA-based cosines was significant [$\chi^2(2) = 106.2$, $p <$

.001]. The coefficients are shown on the right side of Table 3. Seventy-six percent of the cases were correctly classified: 53% and 88% of the minimalist and knowledge-building explanations, respectively. This indicates that without using vector length (or the number of words) as a predictor, 88% of the knowledge-building explanations would be correctly identified. When vector length was included, 86% of all the explanations were correctly classified: 79% and 90% of the minimalist and knowledge-building explanations were correctly classified, respectively.

Overall, the results of the logistic regression equations indicate that the utility of LSA-based predictors alone to predict the use of SERT strategy depends on how one defines noncompliance. If one defines noncompliance as a paraphrase, with a possible extra clause coming from prior text or world knowledge, the equation does well in predicting noncompliance, but not compliance. If one defines noncompliance as only paraphrasing, the equation does an admirable job in predicting whether a student complied— that is, had typed in a knowledge-building response (88% correct). Of course, when vector length is included, the percentage of correct classification increases, indicating its usefulness in classification.

DISCUSSION

This paper presents a novel approach to assessing whether a student is using multiple reading strategies as he or she self-explains a sentence. Developing the procedures for such an assessment is critical for the creation of a computerized version of SERT. Our approach capitalizes on LSA, which computes a measure of semantic analysis between units of language. With LSA, one can simply assess whether the input is more semantically related to words representing the current sentence or words representing the use of particular reading strategies, such as reactivating prior sentences or world knowledge.

Our initial attempts are encouraging for several reasons. First, we verified the assumption that different reading strategies reliably draw upon different sources of information. This is important because instead of attempting to account for an unbounded number of linguistic-based syntactic and semantic cues to assess reading strategy, the computer merely needs to recognize the source of semantic knowledge. Second, the LSA-based assessment of the source of self-explanations was, in fact, remarkably simi-

lar to human-based assessments. This illustrates the validity of the LSA-based approach and is consistent with prior research demonstrating that LSA cosines are similar to human judgments of similarity (Graesser et al., 2000; Landauer & Dumais, 1997; Landauer et al., 1997). Finally, we were able to classify 86% of self-explanations either as minimalist (paraphrases) or as using multiple strategies (knowledge building), with only three LSA-based predictors (i.e., two benchmarks and vector length). Taken together, these findings suggest that LSA will be instrumental in the Web-based practice module of SERT and in similar undertakings.

Despite the encouraging results of the benchmark approach, there is room for improving classification. Of course, one robust predictor was the mean vector length of an explanation. The fact that vector length is correlated with number of words will undoubtedly aid in the classification of self-explanations in the Web-based SERT practice module. However, relying on the number of words in itself will not be sufficient, simply because of the fact that not all long responses indicate the use of multiple strategies. It is also likely that users of a tutor that relies solely on response length to classify strategies use will "catch on" and try to fool the tutor by merely typing in long and, perhaps, incoherent explanations. Classification will also increase if the tutor uses only preselected sentences on which to provide feedback. These sentences will be selected a priori on the basis of the extent to which LSA can distinguish between types of explanations. Finally, the predictive power of the LSA-based variables should increase once we have a dedicated LSA database constructed on the text topics that will be used in the tutor (Shapiro & McNamara, 2000). The LSA database that was used here was based on general reading topics. We are in the process of constructing a database that contains a large sample of texts from life, health, and earth sciences.

Nevertheless, the approach here does little in way of a fine-tuned analysis of the self-explanations. The emphasis here was on whether a reader merely paraphrased the current sentence or was actively engaged in understanding it—a course-tuned analysis. The success of this approach depends on the completeness of the semantic benchmarks. For example, a reader might use an apt metaphor in a self-explanation, but the system will not categorize it as belonging to either prior text or world knowledge if it is novel. Therefore, the current system would not do well with bright and creative individuals when they supply novel explanations. However, that might be the case in many reading assessment tests. Another limitation lies in false world knowledge that readers use during self-explaining. We essentially ignored the topic of incorrect knowledge when the benchmarks were constructed, because we wanted to emphasize whether readers were using world knowledge at all, and not its correctness. One could construct *false information* benchmarks in an attempt to identify these, but we noticed that readers tended not to give incorrect information. Instead, they gave vague or incomplete statements; in fact, it is likely that they chose not to say any-thing, rather than to write down something that they felt could be false.

We should note that we are in the process of evaluating other classification heuristics. One limitation of the present approach is that although the computer can ascertain whether a student is using multiple reading strategies, it cannot ascertain which ones are being used. In order to achieve this goal, we are preparing an exemplar approach to semantic benchmarks. In the exemplar approach, there will be benchmarks representing each of the different reading strategies. For example, a *bridging* benchmark would contain a typical bridge made at that sentence, an *elaboration* benchmark would contain a typical elaboration, and so on. Particular reading strategies would be indicated by the benchmarks with the highest cosines. If this approach proves reliable, SERT feedback can mention particular strategies that are being used by a student and those that are not.

REFERENCES

BIELACZYC, K., PIROLLI, P. L., & BROWN, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition & Instruction, 13*, 221-252.

CHI, M. T. H., DE LEEUW, N., CHIU, M., & LAVANCHER, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

COTÉ, N., & GOLDMAN, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendor & S. R. Goldman (Eds.) *The construction of mental representations during reading* (pp. 169-193). Mahwah, NJ: Erlbaum.

GRAESSER, A. C., WIEMER-HASTINGS, P., WIEMER-HASTINGS, K., HARTER, D., PERSON, N., & THE TRG (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments, 8*, 129-147.

KINTSCH, W. (2001). Predication. *Cognitive Science, 25*, 173-202.

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, 104*, 211-240.

LANDAUER, T. K., LAHAM, D., REHDER, B., & SCHREINER, M. E. (1997). How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society* (pp. 412-417). Mawhwah, NJ: Erlbaum.

MAGLIANO, J. P., & MILLIS, K. K. (2000) *Assessing reading skill with a think-aloud procedure*. Unpublished manuscript.

MAGLIANO, J. P., TRABASSO, T., & GRAESSER, A. C. (1999). Strategic processes during comprehension. *Journal of Educational Psychology, 91*, 615-629.

MCNAMARA, D. S., & SCOTT, J. L. (1999). Training reading strategies. In *Proceedings of the Twenty-first Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

PALINSCAR, A. S., & BROWN, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition & Instruction, 2*, 117-175.

SHAPIRO, A. M., & MCNAMARA, D. S. (2000). The use of latent semantic analysis as a tool for the quantitative assessment of understanding and knowledge. *Journal of Educational Computing Research, 22*, 1-36.

TRABASSO, T., VAN DEN BROEK, P., & SUH, S. (1989). Logical necessity and transitivity of causal relations in the representation of stories. *Discourse Processes, 12*, 1-25.

WIEMER-HASTINGS, P., WIEMER-HASTINGS, K., & GRAESSER, A. C. (1999). Improving an intelligent tutor's comprehension of students with latent semantic analysis. In *Artificial intelligence in education* (pp. 535-542). Amsterdam: IOS Press.

YUILL, N., & OAKHILL J. (1988). Understanding of anaphoric relations

in skilled and less skilled comprehenders. *British Journal of Psychology*, **79**, 173-186.

NOTES

1. LSA is constructed in such a manner that function words have very little impact on LSA cosines.

2. We summed prior text and world knowledge sources because they are conceptually similar, in that both involve adding information to the current sentence from information stored in long-term memory, and because both constitute active uses of SERT strategies. The proportion of prior text (PT) and world knowledge (WK) clauses was lower under a minimal reading strategy (PT, M = .01; WK, M = .01) than under a sentence-focused strategy (PT, M = .15; WK, M = .26), which was, in turn, lower than under a knowledge building strategy (PT, M = .23; WK, M = .38).

The pattern was similar to the corresponding LSA values. The proportion of PT and WK clauses was lower under a minimal reading strategy (PT, M = .23; WK, M = .18) than under a sentence-focused strategy (PT, M = .28; WK, M = .20), which was, in turn, lower than under a knowledge building strategy (PT, M = .34; WK, M = .22). However, it is evident that differences in overlap with prior text benchmarks carried more weight in the reported differences across the strategies than did the world knowledge benchmarks. This may be due to the fact that explanations are considerably more constrained by the prior text than by world knowledge. As such, it is more difficult to identify possible explanations based on world knowledge than those based on prior text.

3. Replacing vector length with the number of words increased the Nagelkerke R^2 from .53 to .66 and from .69 to .80 for the minimalist/sentence-focused versus knowledge-building and the minimalist versus knowledge-building analyses, respectively. Therefore, one practical way to improve classification is to use the number of words, rather than vector length.

APPENDIX

Heart Disease
1. The heart is the hardest-working organ in the body.
2. We rely on a regular blood supply every moment of every day.
3. Any disorder that stops the blood supply is a threat to life.
4. More people are killed every year in the U.S. by heart disease than by any other disease.
5. A congenital disease is one with which a person is born.
6. Most babies are born with perfect hearts.
7. In about one in every 200 cases something goes wrong.
8. Sometimes a valve develops the wrong shape.
9. It may be too tight, or fail to close properly.
10. Sometimes a gap is left in the septal wall between the two sides of the heart.
11. When a baby's heart is badly formed, it cannot work efficiently.
12. The blood does not receive enough oxygen.
13. The blood cannot get rid of carbon dioxide through the lungs.
14. It becomes purplish, and the baby's skin looks blue.
15. The baby is in danger of suffocating.
16. Diseases can sometimes cause the heart to not form properly.
17. The disease called rheumatic fever may cause harm to the heart.
18. The disease usually follows a sore throat caused by bacteria called streptococci.
19. The tissues of the heart become inflamed.
20. If it is badly affected, it fails.
21. Usually it recovers, and the results of the damage are seen only years later.
22. The valves of the heart are left with scars.
23. They cannot work properly.
24. Eventually it may fail.
25. The effects of the rheumatic fever may take up to twenty or thirty years to appear.
26. The most common heart problem that we think about is a heart-attack.
27. The blood vessels that extend across the heart and supply it with blood are called the coronary arteries.
28. They give the heart the oxygen it needs to carry on working.
29. If they become blocked, parts of the heart muscle will die.
30. This causes the patient to have a heart attack, which can be fatal.
31. The blockage of a coronary artery is usually caused by a thrombus, or blood clot.
32. Coronary thrombosis happens when a clot forms in a coronary artery.
33. That is the correct name for a heart attack.
34. Whether heart disease is congenital, caused by other diseases, or the result of a blood clot in the coronary arteries, it is a very serious problem that requires medical attention.