Choosing Reading Passages for Vocabulary Learning by Topic to Increase Intrinsic Motivation

Michael HEILMAN+, Alan JUFFS*, Maxine ESKENAZI+
+Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15221 U.S.A
*Department of Linguistics, University of Pittsburgh
Pittsburgh, PA 15260 U.S.A.

Abstract: Intrinsic motivation has been shown in previous research to lead to better learning. In order to increase intrinsic motivation, REAP, a tutoring system for ESL vocabulary was enhanced to prefer practice readings that match personal interests. In a randomized experiment, students receiving personalized readings indicated higher levels of interest in post-reading questionnaires. Additionally, overall post-test scores were somewhat higher (but not significantly) for students with interest-matched practice readings than for students using a previous version of REAP that did not match topics to student interests.

1. Introduction

Intrinsic motivation is the desire to learn about something for its own sake. In contrast, extrinsic motivation is the desire to learn in order to accomplish a specific task, such as a practice problem or an assessment. Intrinsically motivated students are more likely to take risks, choose difficult learning paths, persist in the face of difficulty, and apply effective learning strategies [1]. Interested—and thus more intrinsically motivated—students are also more likely to deeply process material. Studies have found that students who are more interested in readings show evidence of deeper processing. Interested students performed better on complex questions about a reading, as well as questions that require them to apply knowledge to novel situations [2].

It seems desirable, therefore, to increase student interest in order to promote intrinsic motivation, deeper processing of material, and better learning strategies. Work by Cordova and Lepper [3] showed positive effects of contextualization, personalization, and choice within an educational game for children in the domain of arithmetic. In some cases, however, attempts to increase interest may adversely affect progress through a given curriculum. de Soldato and du Boulay [4] discussed tradeoffs between a tutor’s motivational plan and its domain-based (i.e., instructional or pedagogical) plan.

This paper discusses the enhancement of the REAP tutor to allow for personalization of reading materials by topic in order to increase interest and motivation. In this work, the term “personalization” refers to the selection of practice readings in order to match a student’s interests. The REAP tutor is an intelligent tutoring system for English as a Second Language (ESL) vocabulary and reading practice [5]. The tutor provides contextualized practice on individualized vocabulary lists by selecting reading passages (roughly 1000 words long) from a large corpus of annotated Web documents. There are a variety of constraints that the tutor considers when selecting readings for students, including reading difficulty level, grammaticality, scheduling of practice, the length of a reading, the number of target words in a reading, etc. The tutor selects reading materials from its corpus that contain target words from individualized lists and satisfy these other constraints.

During each training session, students work through a series of readings, each of which is followed by practice exercises for the target words in the reading. While reading a passage, students are able to access dictionary definitions for any word in a reading either by clicking on a highlighted target word or by typing a word into a box in the lower-left corner of the screen. The target words in the readings are also highlighted because highlighting may increase the use of dictionary definitions [6], thus encouraging students to coordinate multiple sources of information about a word’s meaning—namely, the implicit context around words and the explicit definitions of words.

A problem discovered in past studies with REAP is that many students spend only a brief amount of time on a reading and do not deeply process the text. Students often only read the dictionary definition for target words rather than attempting to process the entire context around the words. Inferring the meaning of vocabulary from context is a seemingly important strategy that is not used by such students (see [7] for a review of incidental vocabulary acquisition). This behavior is likely due to a desire to perform well on post-reading practice exercises and post-test, which can be viewed as forms of extrinsic motivation. Intrinsically motivated students who are more interested in a reading are more likely to read the entire text and to use...
context to learn the meaning of unknown vocabulary. Therefore, personalization that increases intrinsic motivation could lead to deeper processing of context and better learning of vocabulary.

2. Text Classification for Personalization of Reading Material

To allow for the personalization of readings, the REAP tutor includes personalization by topic as a factor in its algorithm for choosing optimal readings. Students take a short survey after the pre-test to inform the system about which general topics they are interested in reading about. The system then prefers readings that have been classified as pertaining to those topics. Students do not receive readings, however, containing only narrow-coverage words that relate only to their topics of interest. For example, students interested in sports should not get readings with target words such as “football.” Students with different interests practiced similar sets of general purpose vocabulary. For instance, in the study described in this paper, one student interested in arts saw the word “endure” in a text describing an artist’s early career struggles (“For an artist who has endured so many years of obscurity...”). Another student interested in business saw the same word used to describe economic hardship (“As California has endured a burst tech bubble, costly energy crisis and a staggering burden on its business community...”).

In order to improve interest levels of students, a text classification system was implemented to classify each potential reading by its general topic. A Support Vector Machine [8] text classifier with a linear kernel was trained on Web pages from the Open Directory Project (ODP, http://dmoz.org), which are organized into a hierarchy of topics. SVM-Light [9] was used as the implementation of Support Vector Machines. The following general topics were manually selected from the set of top-level ODP categories: Arts, Business, Computers, Games, Health, Home, Recreation, Science, Society, and Sports. Web pages with human-assigned topic labels from the ODP (1,000 pages/topic) were used as training data for the classifier. These pages were first filtered so that their length was similar to that of the readings in the corpora used by the tutor.

Two quantitative evaluations were conducted before the deployment of the text classification system to ensure that its labels were sufficiently accurate (some details are omitted for brevity). First, the binary classifiers for each topic category were evaluated according to precision, recall, and the F1 measure—the harmonic mean of precision and recall. For the evaluations of each classifier, leave-one-out cross validation was performed using SVM-Light. The binary classifiers had macro-average (average across categories) precision, recall and F1 measures of 0.78, 0.74, and 0.77, respectively.

Post-reading interest questionnaire results indicate that the topic choice system in REAP is effective at improving interest. After each reading, students were asked how interesting the just-completed reading was on a Likert scale from one to five, with five indicating greatest interest. Students in the treatment condition (described below) with personalization of readings by topic responded that they were interested in readings more frequently than did students in the control condition. The distribution of responses is shown in Table 1.

### Table 2: Post-reading interest responses for students using versions of tutor with or without personalization of readings by topic.

<table>
<thead>
<tr>
<th>Score</th>
<th>Percentage of Scores (Without Personalization)</th>
<th>Percentage of Scores (With Personalization)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (least interest)</td>
<td>6.9%</td>
<td>4.1%</td>
</tr>
<tr>
<td>2</td>
<td>15.0%</td>
<td>19.1%</td>
</tr>
<tr>
<td>3</td>
<td>41.9%</td>
<td>32.0%</td>
</tr>
<tr>
<td>4</td>
<td>30.0%</td>
<td>39.7%</td>
</tr>
<tr>
<td>5 (most interest)</td>
<td>6.2%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

3. Experimental Design

An experiment was conducted to measure the effects of personalization on learning progress in the REAP tutor. Thirty-five students at the English Language Institute at the University of Pittsburgh participated in this experiment as part of an intermediate English as a Second Language Reading course. The students were
randomly assigned to control or treatment conditions. Of the students who completed the experiment, 19 were in the control condition and 16 in the treatment condition. For students in the control condition, the REAP tutor ignored student interest survey results and offered readings to students based on the goals of the curriculum. For students in the treatment condition, the REAP tutor was the same except that it also preferred readings about topics of personal interest.

At the start of the study, each student was asked to perform self-assessments to create an individualized curriculum consisting of a set of words from the Academic Word List [10] that he or she identified as unknown. There were nine training sessions of 40 minutes, and one post-test session. During training, each student worked through a series of practice readings and exercises with the REAP tutor.

At the end of the series of training sessions, students took a post-test consisting of cloze questions for the target vocabulary words that were identified as unknown through self-assessments. For more information about the use of cloze questions in the REAP tutor, see [11]. The post-test consisted of forty questions for target words that appeared in at least one passage completed by the student. For students who did not practice forty words, additional questions for unpracticed target words were added.

4. Results and Conclusion

The effect of personalization on learning was measured by student performance on the post-test cloze questions for target vocabulary words. The results are shown in Figure 1. Students in the treatment condition performed better on average (M=35.5%, SD=14.9%) in terms of overall post-test scores compared to students in the control condition (M=27.1%, SD=17.2%). The difference in mean overall post-test scores in the treatment condition was 8.4% (95% CI = -2.8%, 19.5%), which corresponds to a medium effect size of 0.51. However, this difference is not statistically significant (p=0.14, two-tailed t-test).

The tutor was not able to provide practice for all words on the post-test for all of the students, due to a lack of practice texts with multiple words and the slow pace of some students. Therefore, performance for only practiced words was also analyzed. Students in the treatment condition correctly answered a higher percentage of post-test questions for words practiced in at least one reading (N=16, M=50.3, SD=20.1) than did students in the control condition (N=19, M=32.4, SD=18.9). A two-tailed t-test for independent means verified that this result is statistically significant (t=2.719, df=33, p=0.005). The difference of scores between the two groups was 17.9% (95% CI = 4.5%, 31.3%), which corresponds to a large effect size of 0.85. This result indicates that personalization improved learning for the words that students had the opportunity to practice in at least one reading. This work suggests that automatic techniques can be effectively applied to select readings by topic to match student interests. The results for measures of learning are promising and suggest that the effects of personalization of texts for vocabulary practice should be investigated further.

5. Acknowledgements

The authors would like to acknowledge Kevyn Collins-Thompson, Jamie Callan, and Lois Wilson for their work on the REAP project. Thanks also go to the anonymous reviewers. This material is based on work supported by NSF grant IIS-0096139. Any opinions, findings, conclusions or recommendations expressed in this material are the authors', and do not necessarily reflect those of the sponsor.

References