Diagnosing Reading Deficiencies of Adults with Low Literacy Skills in an Intelligent Tutoring System

Genghu Shi[1,2], Andrew J. Hampton[1,2], Su Chen[1,2], Ying Fang[1,2], and Arthur C. Graesser[1,2]

[1] University of Memphis, Memphis TN 38111, USA
[2] Institute for Intelligent Systems
gshi@memphis.edu

Abstract. We developed a version of AutoTutor that helps struggling adult learners improve their comprehension strategies through conversational agents. We hypothesized that the accuracy and time to answer questions during the conversation could be diagnostic of their mastery of different reading comprehension components: words, textbase, situation model, and rhetorical structure. The results show that adults' performance on more basic reading components (i.e., meaning of words) was higher than on the deeper discourse levels. In contrast, time did not vary significantly among the theoretical levels. The results suggested that adults with low literacy had higher mastery on basic reading levels than deeper discourse levels. The tracking of performance on the four theoretical levels can provide a more nuanced diagnosis of reading problems than a single overall performance score and ultimately improve the adaptivity of an ITS like AutoTutor.

Keywords: CSAL AutoTutor, Reading strategies, Comprehension framework.

1 Introduction

We developed a version of a web-based intelligent tutoring system (AutoTutor) for adults with low literacy skills to improve their reading comprehension strategies in the Center for the Study of Adult Literacy (CSAL). AutoTutor for CSAL has 35 lessons that focus on distinct theoretical levels of reading comprehension articulated by Graesser and McNamara [1]. For each lesson, the system starts out assigning words or texts at a medium difficulty level and then asks 8 to 12 multi-choice questions about them. In this study, we tracked four theoretical levels (of the six defined in [1]). *Word* represents the lower-level basic reading components. The other three theoretical levels (*textbase, situation model*, and *rhetorical structure*) represent deeper discourse levels. We hypothesized that the accuracy and time on questions in AutoTutor could be diagnostic of adults' mastery of comprehension components. Therefore, by comparing the accuracy and time on questions of four theoretical levels, we can detect adults' strengths and weaknesses in reading competencies.

Cite as

Shi, G., Hampton, A., Chen, S., Fang, Y. and Graesser, A. (2018). Diagnosing reading deficiencies of adults with low literacy skills in an intelligent tutoring system. In R. Nkambou, R. Azevedo & J. Vassileva (Eds.), *Proceedings of the 14th International Conference on Intelligent Tutoring Systems* (pp 463-465). Montreal, Cananda: Springer.

2 Methods

2.1 Participants

The participants were 52 adults recruited from CSAL literacy classes in Atlanta and Toronto. They completed a 100-hour intervention over four months. Their ages ranged from 16–69 years (Mean = 40, SD = 14.97). The majority of the participants were female (73.1%). All participants read at 3.0–7.9 grade levels.

3 Measures and Data Analysis

We extracted the adults' initial responses on medium level questions in each of the 29 lessons that focused on the four theoretical levels. All adults answered these initial medium questions before adaptively branching to easy or difficult questions in AutoTutor. The initial responses included accuracy (1 or 0) and time to select an answer (in seconds).

We performed a descriptive analysis by exploring the means and standard deviations of accuracy and time on questions of the four theoretical levels. Then we performed mixed effect models [2] on the two measures to test the difference among the four theoretical levels, with *question* as the unit of analysis. The random effects were participants, lessons, and questions; the fixed effect was theoretical level. Participants' random slopes on different theoretical levels and random intercepts of the interaction between lesson and question were also included in the models.

4 Results

Table 1 shows the means of accuracy and time on questions separately as a function of the four theoretical levels. The pattern of scores indicate that performance is highest and answer times are shortest for the *word* level (reference level in the analysis) compared to the three discourse levels (*textbase, situation model,* and *rhetorical structure*).

A Type II Wald Chi-square test on the logistic mixed effect model showed that accuracies were significantly different ($\chi^2(3) = 8.34$, $p = 0.04$) among the four theoretical levels. A post-hoc analysis with pairwise comparison showed only *word* pairs were significantly different. An ANOVA of type III with Satterthwaite on linear mixed effect model showed that time did not vary among the four theoretical levels, $F(3, 25.8) = 0.058$, $p = 0.981$.

Table 1. Means and Standard Deviations of Accuracies and Time

		Word	Textbase	Situation Model	Rhetorical Structure
	No. of Questions	1455	1981	5049	5071
Accuracy	Mean (SD)	0.80 (0.40)	0.69 (0.46)	0.67 (0.47)	0.69 (0.46)
Time	Mean (SD)	31.7 (30.4)	35.1 (30.2)	35.2 (31.6)	37.1 (38.1)

5   Discussion and Conclusion

The logistic mixed effect model indicates that adults' performance on *word* level was higher than the three discourse levels. This likely occurred because word items focused on individual words or single sentences which require low loads on working memory, whereas solving the items of deeper discourse levels is time-consuming, strategic, and taxing on cognitive resources. The time that adults spent on questions were not significantly different across theoretical levels, although times trended slower as theoretical levels progressed.

This study provides a more nuanced diagnosis of adults' reading problems within a multilevel reading comprehension framework than a single overall performance score could contribute. Future research should focus on designing standard reading tests and establishing norms for adult populations based on the multilevel framework that affords this diagnostically useful differentiation. Combining the testing results and the norm, researchers could develop more adaptive intelligent tutoring systems which provide customized learning contents to low literacy adults.

6   Acknowledgements

References

1. Graesser AC, McNamara DS (2011) Computational analyses of multilevel discourse comprehension. Topics in Cognitive Science, 3 (2), 371-398
2. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:1–48