When do Students Interrupt Help? Effects of Time, Help Type, and Individual Differences

Cecily HEINER, Joseph BECK, and Jack MOSTOW
Carnegie Mellon University
Project LISTEN, Newell Simon Hall, 5000 Forbes Avenue Pittsburgh, PA 15213

Abstract. When do students interrupt help to request different help? To study this question, we embedded a within-subject experiment in the 2003-2004 version of Project LISTEN's Reading Tutor. We analyze 168,983 trials of this experiment, randomized by help type, and report patterns in when students choose to interrupt help. Using the amount of prior help, we fit an exponential curve to predict interruption rate with an r^2 of 0.97 on aggregate data and an r^2 of 0.22 on individual data. To improve the model fit for individual data, we adjust our model to account for different types of help and individual differences. Finally, we report small but significant correlations between a student parameter in our model and external measures of motivation and academic performance.

1. Introduction

One of the key benefits of an intelligent tutoring system is on-demand help. However, students do not always appreciate the help that a tutoring system gives, and students may interrupt some forms of help in order to receive more desirable help. In this paper, we study when students interrupt help and we report results which may assist in designing better research and better help systems.

We describe work that relates interruption in a user-system dialogue, help seeking, and user modeling. Earlier research has reported on a system that interrupts when a student deviates from correct behavior, hesitates, or gets stuck[1], but interrupting a user has a cost [2]. When students interrupt the system, they are often seeking help, a behavior that can be predicted with 83% accuracy [3]. However, researchers have found that students often do not know when they need help[4], and they have identified help abuse as a problem that accounts for a third or more of help seeking bugs[5]. One particular example of help abuse is described as "gaming the system"[6], a pattern in which some students "click through" help until they receive the answer with negative consequences to their learning. Extending the idea that help-seeking behaviors relate to the attributes of a specific student, researchers have inferred student variables based partially on help-seeking and help-usage behavior[7]. However, a more integrated approach to understanding help-seeking behavior and user-modeling could improve both our student models and our assessments of what they know[8]. In this paper, we present an experiment and series of analyses that examines interruption and user-system dialogue, help seeking behavior, and user modeling together.

In this paper, we present an experiment, first explaining the design behind the experiment and then describing the data set. Next we analyze the results of that experiment and use them to fit a series of models. Then we interpret the results and correlate them with other data. Finally, we conclude and suggest future work.

2. Experimental Design

Our data come from the 2003-2004 version of the Project LISTEN Reading Tutor, which presented text and used automatic speech recognition to listen to children read aloud[9]. When a student encountered a difficult word, the student could click on that word for help. Alternatively, the Reading Tutor may have detected that a student was struggling with a word and taken the initiative to give spoken or graphical help. Regardless of whether the help was student initiated or tutor initiated, the student could interrupt that help and receive more help by clicking on the word again while the Reading Tutor was speaking. Sometimes, students clicked two or more times, interrupting the previous help with each click and triggering a new instance of help. Alternately, a student may have clicked on a word for help more than once, but waited until the previous help had completed before clicking again. Each time the Reading Tutor gave help, it chose randomly from a variety of applicable and efficacious help types without regard to previous help that it had given. The Reading Tutor primarily gave the nine types of word help listed below[10].

- ***SayWord*** plays a recording of the word. e.g. "cat"
- ***WordinContext*** plays a recording of the word extracted from the sentence.e.g. "…cat…"
- ***Autophonics*** pronounces a selected grapheme in the word. e.g. "c here makes the sound /k/"
- ***SoundOut*** plays video clips of a child's mouth saying the phonemes of the word. e.g. "/k/…/ae/…/t/"
- ***Recue*** reads words in the sentence leading up to, but not including, the word. e.g. "I have a dog and a"
- ***OnsetRime*** says the first phoneme, pauses, and says the rest of the phonemes. e.g. "/k/…/ae/ /t/"
- ***StartsLike*** says "starts like (word with the same beginning)." e.g. "starts like cats"
- ***RhymesWith*** says "Rhymes with (rhyming word)." e.g. "rhymes with mat"
- ***Syllabify*** says the syllables of the word separated by short pauses. e.g. "cat"

The Reading Tutor randomized the choice to provide a variety of help [1] and to embed an experiment to compare the effectiveness of different types of help[11].

This embedded experiment examines when students interrupt help. Each randomized trial starts when the student or tutor initiates help. The randomized variable is the selection of help type selection. Another analysis[10] considered students' subsequent performance when reading the word as the outcome variable. In this experiment, the outcome variable is whether or not students interrupt help. The experiment is diagrammed in Figure 1.

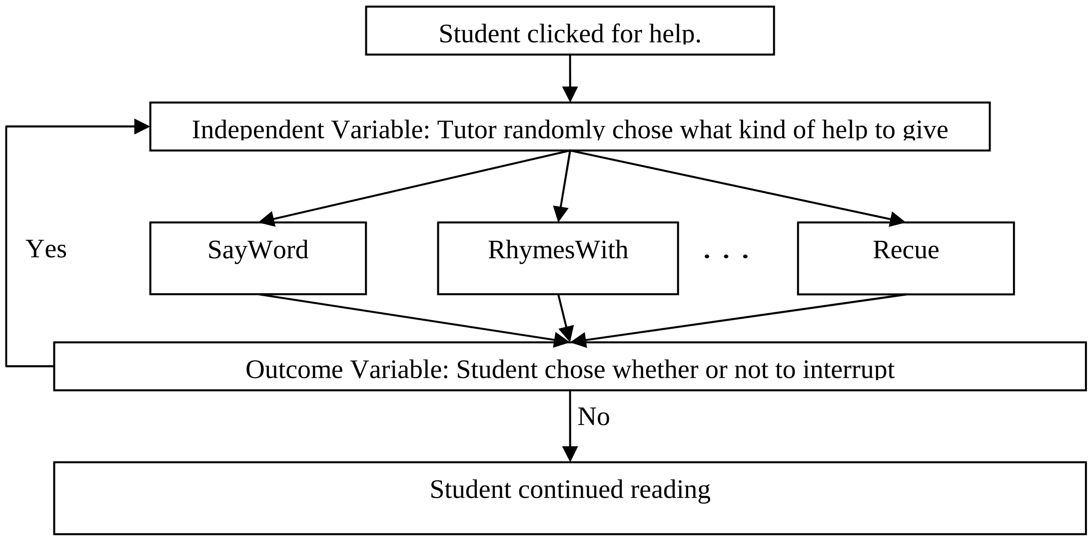

Figure 1: Experiment Diagram

3. Modeling Approach

We considered prior exposures to help as an important independent variable in our experiments. In this paper, we define prior exposures to help as the number of times a particular user has received help of a particular type before. Because not all students receive the same amount of exposure to the Reading Tutor, we were concerned that maybe some students interrupted more because they had more exposure to the Reading Tutor; in other words we thought maybe we were seeing an effect of attrition. To insure that our trend could not be explained by attrition, we included the first hundred trials for a given student and help type, excluding students and help types with fewer. We chose 100 because Figure 2 showed that it was large enough to reveal an asymptote, but not so large that it would eliminate data unnecessarily or create a bias favoring certain models. This step left data from 368 K-4 students with a variety of reading abilities. We did not distinguish between tutor-initiated (15% of the data) and student-initiated (85% of the data) help or exclude students who had not received all nine help types from this data set.

3.1 Fixed Parameter Model

Plots of the data in Figure 2 show the exponential relationship in Equation 1. To estimate the values of the parameter a, b, and c of the equation accurately, we used SPSS[12], software for statistical analysis, to do a non-linear regression analysis. We used a Java applet [13] to initial parameter estimates for a curve with a shape similar to our data. Using the initial parameters, SPSS found that a= -0.27, b= -0.08, and c=0.45 in this model. After considering other models including a power curve and a logarithmic model, we selected the exponential curve because other researchers have suggested that it is a better fit for individual data [14], and it fit our data best.

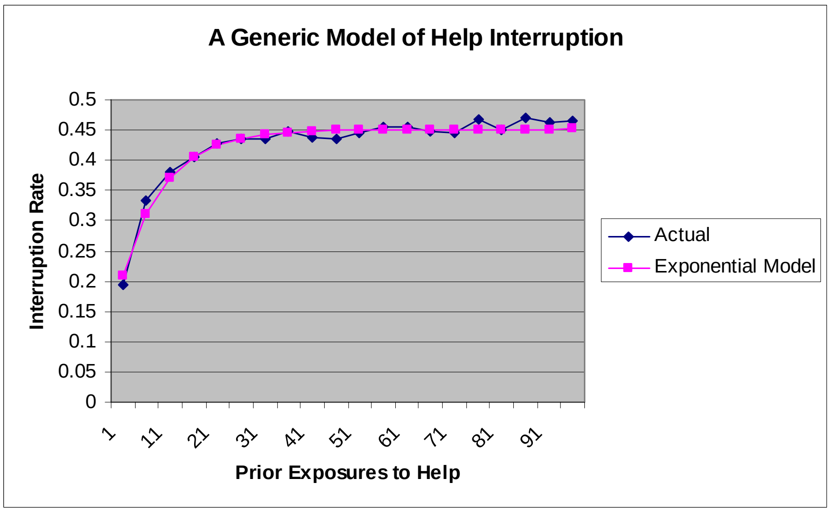

Figure 2: Simple Model

$$P(i) = a * (e^{b * prior\_exposures}) + c$$

Equation 1: Model Form

As shown in Figure 2, the actual data and the exponential model are correlated. The r^2 of .97 shows that the aggregate data fit the model well. We speculated that this equation may be a learning curve for recognizing undesirable help. We also speculated that the parameters in this equation are related to properties of a specific help type or the student. We caution that this model is fit using aggregate data; a model fit to an individual student's data with a binary outcome variable will have a much lower r^2, 0.22 for this dataset. For both the aggregate and the individual data, the asymptote represents the interruption rate after students have developed habits that alter how they use the system. Because interruption rates vary by help type, we believed we could add conceptual value to the model and improve its fit for individual data by adding parameters to account for help type and individual differences.

3.2 Fitting a Model with Help Type Parameters

To build a model that was more closely related to help type, we adjusted parameters *a*, *b*, and *c* by refitting them for each help type, instead of aggregating all the help types together. To estimate the parameters of the equation for each help type, we did another SPSS non-linear regression. Using generic values as initial estimates of the parameters, SPSS found the values in Table 1.

Table 1: Help Types

Help Type	*a*	*b*	*c*	Interruption Rate	Duration
SayWord	-.11	-.06	.14	.12	.69 s
WordInContext	-.10	-.21	.19	.18	1.11 s
OnsetRime	-.29	-.07	.39	.35	1.19 s
SoundOut	-.41	-.06	.55	.47	2.91 s
Syllabify	-.39	-.15	.54	.52	2.00 s
RhymesWith	-.33	-.08	.59	.55	2.36 s
StartsLike	-.38	-.11	.61	.58	2.60 s
Autophonics	-.49	-.09	.75	.69	2.43 s
Recue	-.31	-.07	.79	.75	3.89 s

Now we had nine values for each parameter and we could look for patterns. When we correlated our *c* parameter with the interruption rate for a help type we found an r^2 of .99. Independent of the model, another clear pattern relates the interruption rate for a particular help type and duration with an r^2 of .79. Thus the *c* parameter models the average interruption rate for a particular type of help which is related to the duration of help. This makes sense since students have more time to move their mouse and click to interrupt when help has a longer duration. We did not find clear patterns for *a* and *b*, but we hypothesize that that *a* and *b* may related to how many exposures a student needs to learn to recognize a specific help type and interrupt it because these parameters are lower for the help types that give the answer; we do not currently have the data and analysis to confirm this idea.

3.3 Fitting a Model with a Student Parameter

To better account for individual differences, we added one more parameter to our equation. We held the help type parameters of the equation at their values from the previous regression and estimated a new student parameter *s*, applying SPSS non-linear regression to Equation 2. Conceptually, this student parameter, *s*, alters the asymptote of the graph and is related to a student's interruption rate, a value that should be between zero and one. To insure that values for *s* would be consistent with this idea, we altered the form of the model slightly, setting the initial value for the student parameter *s* at 1, and imposing the limits that *s* must be less than or equal to 1 and greater than -.5. Within this range [-.5, 1], SPSS fit a single student parameter for each student.

$$P(i) = a * (e^{b*opportunities}) + c + (\mathrm{s} * (1 - c))$$

Equation 2: Student Parameter Model

3.4 Evaluating the Relative Value of the Various Models

Table 2 compares the various models and two additional baseline models, using mean squared errors and r^2. The overall interruption rate model simply predicts that 43% of all help will be interrupted, since this is the average interruption rate when all of the data is aggregated together. The mean interruption by help type model predicts help interruption based on the interruption rate for a given help type. We included both of these baselines to measure how much variance the help type accounts for on its own. Table 2 shows that the biggest reductions in mean square error and improvements in r^2 come from applying a generic model that takes time into consideration by accounting for the amount of previous help. Fitting the model based on help type improved the model a little, but not much. Adding a student parameter improved the model moderately.

Table 2: Models and Mean Square Errors

Model Name	Mean Square Error	r^2
Overall Interruption Rate	.24	-
Mean Interruption by Help Type	.24	0.01
Generic Model with Prior Help	.19	0.22
Help Type Parameters	.19	0.24
Student / Help Type Parameters	.17	0.30

4. Correlating the Student Parameter against External Measures

The student parameter, *s* in the final model is a variable that may relate to other measures of a student, including process variables and test scores. We considered the following process variables: help request rate, help interruption rate, disengagement (measured as the percentage of questions that students answer hastily[15]), and percentage of time picking stories. We were surprised that we did not find correlations with other affective variables such as disengagement or help request rate.

For test scores, we considered pre- and post-test scores and gains for the Elementary Reading Attitude Survey (ERAS) [16] and a fluency test. ERAS is a twenty item instrument with ten items each for recreational and academic reading attitudes. The fluency test consists of a timed, levelled reading passage which students read for trained fluency testers. Small, significant negative correlations exist between ERAS academic and motivational test scores. So, *s* relates to attitudes towards academic and recretaional reading. Additionally, small but insignificant correlation exists between fluency pre-test and the student parameter *s*. So, *s* may also be related to fluency. Table 3 displays the meaningful correlations.

Table 3: Student Parameter Correlations

Test Name	Pearson Correlation	Signficance
Fluency Pre-Test	-.155	.072
ERAS Recreational Pre-Test	-.267	.002
ERAS Academic Pre-Test	-.283	.001

In order to determine the relationship between *s* and gender, we ran an independent T-test and found the mean *s* value for girls is -0.057 and the mean *s* value for boys is 0.037 with a p-value of <0.001. This means that girls are less likely to interrupt than boys, the difference is significant, and the *s* parameter is related to gender.

Future Work and Conclusion

This paper is the first to study when students interrupt spoken help and to propose a predictive model of this behavior. An exponential model characterizes the temporal aspect of this behavior and shows that the number of previous exposures to a particular type of help is an important predictor of whether or not a student will interrupt help. We report values for three parameters that characterize help type and show that one of them correlates highly with the interruption rate for a given help type. Additionally, girls are less likely to interrupt than boys. Interruption rates are somewhat negatively correlated with pre-test scores, so less motivated poor readers interrupt more. We compare successively refined models for predicting help interruption rate; the biggest improvement in model fit comes from accounting for temporal factors. The exponential model could be a learning curve for recognizing help that students find undesirable.

This paper has illuminated how students use help. We have suggested that there is an initial window of adaptation when students are learning to use an intelligent tutoring system. After this window, students interrupt each help type at an approximately constant rate. In our model, the interruption rate approaches the asymptote when the student has had approximately thirty prior encounters with a particular kind of help. Thirty prior encounters of help roughly corresponds to an average of three hours of system usage spread across eighteen sessions or six weeks of calendar time. These patterns suggest a need for long-term studies to understand how students use intelligent tutoring systems after they have adapted to them. We have also proposed that initial data should be considered separately due to startup effects.

This paper is one step towards the long-term goal of being able to quantify affective factors and link them to learning gains. We still do not know very much about why students interrupt help. Are they bored, tired, lazy, impatient, or rude? What are students looking for when they interrupt help? The answers to these questions might suggest how we can encourage students to tolerate long, laborious, but educational help.

Acknowledgements

This work was supported by the National Science Foundation, under ITR/IERI Grant No. REC-0326153. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the official policies, either expressed or implied, of the sponsors or of the United States Government.

Thanks to Ryan Baker, Hao Cen, and Jonathan Brown for comments on drafts of this paper. Additional thanks to students, teachers, and administrators in schools where the Reading Tutor is deployed.

References (website: www.cs.cmu.edu/~listen/)

1. Mostow, J. and G. Aist, *Giving help and praise in a reading tutor with imperfect listening -- because automated speech recognition means never being able to say you're certain.* CALICO Journal, 1999. **16**(3): p. 407-424.
2. Horvitz, E. and J. Apacible. *Learning and Reasoning about Interruption*. in *ACM International Conference on Multimodal Interfaces*. 2003. Vancouver, Canada.
3. Beck, J.E., P. Jia, J. Sison, and J. Mostow. *Predicting student help-request behavior in an intelligent tutor for reading*. in *Proceedings of the 9th International Conference on User Modeling*. 2003. Johnstown, PA.
4. Aleven, V. and K.R. Koedinger. *Limitations of Student Control: Do Student Know when they need help?* in *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*. 2000. Montreal: Berlin: Springer Verlag.
5. Aleven, V., B. McLaren, I. Roll, and K.R. Koedinger. *Toward tutoring help seeking Applying cognitive modeling to meta-cognitive skills*. in *Intelligent Tutoring Systems*. 2004. Maceio, Alagoas, Brazil: Springer.
6. Baker, R.S., A.T. Corbett, K.R. Koedinger, and A.Z. Wagner, *Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System"*. Proceedings of ACM CHI, 2004: p. 383-390.
7. Arroyo, I., T. Murray, and B.P. Woolf. *Inferring observable learning variables from students' help seeking behavior*. in *Intelligent Tutoring Systems. Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes*. 2004. Maceio, Brazil.
8. Wood, H. and D. Wood, *Help seeking, learning, and contingent tutoring.* Computers & Education, 1999. **33**(2-3): p. 153-169.
9. Mostow, J. and G. Aist, *Evaluating tutors that listen: An overview of Project LISTEN*, in *Smart Machines in Education*, K. Forbus and P. Feltovich, Editors. 2001, MIT/AAAI Press: Menlo Park, CA. p. 169-234.
10. Heiner, C., J.E. Beck, and J. Mostow. *Improving the Help Selection Policy in a Reading Tutor that Listens*. in *Proceedings of the InSTIL/ICALL Symposium on NLP and Speech Technologies in Advanced Language Learning Systems*. 2004. Venice, Italy.
11. Heiner, C., J. Beck, and J. Mostow. *Lessons on using ITS data to answer educational research questions*. in *Proceedings of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes*. 2004. Maceio, Brazil.
12. SPSS, *SPSS for Windows*. 2000, SPSS Inc.: Chicago, IL.

13. Dendane, A., *Experiment and Explore Mathematics: Tutorials and Problems*. 2005.
14. Heathcote, A., S. Brown, and D.J.K. Mewhort, *The Power Law Repealed:The Case for an Exponential Law of Practice.* Psychonomics Bulletin Review, 2000: p. 185-207.
15. Beck, J.E. *Using response times to model student disengagement*. in *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*. 2004. Maceio, Brazil.
16. Kush, J.C. and M.W. Watkins, *Long-term stability of children's attitudes toward reading.* The Journal of Educational Research, 1996. **89**: p. 315-319.