Does learner control affect learning?

Joseph E. BECK
Machine Learning Department, Carnegie Mellon University
joseph.beck@gmail.com

Abstract. Many intelligent tutoring systems permit some degree of learner control. A natural question is whether the increased student engagement and motivation such control provides results in additional student learning. This paper uses a novel approach, learning decomposition, to investigate whether students do in fact learn more from a story they select to read than from a story the tutor selects for them. By analyzing 346 students reading approximately 6.9 million words, we have found that students learn approximately 25% more in stories they choose to read, even though from a purely pedagogical standpoint such stories may not be as appropriate as those chosen by the computer. Furthermore, we found that (for our instantiation of learner control) younger students may derive less benefit from learner control than older students, and girls derive less benefit than boys.

1 Introduction

Many intelligent tutoring systems (ITS) have some element of learner control. For example, the Reading Tutor [1] sometimes allows students to select which story they will read next. Furthermore, most tutoring systems allow students to determine when to ask for help (e.g. [2] and see [3]). Learner control is certainly helpful in making students feel more engaged with their learning, and engaged learners tend to learn more [4]. Presumably, the transitive argument holds that allowing learner control will result in more student learning. However, there are possible drawbacks to learner control. First, by their very nature students are not expert in the content to be learned. While it might make sense to allow a thirty-five year old who is using a computer based training system to study for an exam for promotion to control his own learning, it is far less obvious that allowing a six year old such control will be as productive. Students do not always make optimal choices for their learning [3]. The notion of "perceived learner control" [5] avoids these problems. In this framework, the choices students are given do not have large implications for the actual domain content seen. For example, students could choose which icon will be used to represent progress (e.g. [6]). The second drawback of learner control is that computers can make decisions much more quickly than can students. Time spent letting the student make decisions is time spent not learning the material. Given the primacy of time on task for learning, this trade should be made with extreme caution.

This paper addresses the following questions: Does perceived learner control influence how much students learn? By what amount? Which students benefit? Are any gains simply from novelty or are they persistent? And does the benefit from allowing students to influence their learning outweigh the time lost from learning activities that such interactions entail?

2 Approach

In Project LISTEN's Reading Tutor [1], the student and tutor take turns picking which story the student is going to read next. As shown in Figure 1, when it is the student's turn to select a story, the tutor presents four stories and how many times the student has read that story previously. The student then clicks on whichever story he wishes to read. If the student does not like the four stories presented, he can select to see other stories at the same level of difficulty, or he can select stories at a different level of difficulty. To avoid the stigma of being behind one's peers, the story levels are somewhat awkwardly encoded: "K" represents the lowest story level (short for "kindergarten"), followed by "A," then "B," up to level "G." Although the interface may seem somewhat baroque, finding a design that works for young children who are barely able to read is challenging. See [7] for more details of the interface for selecting stories. When it is the Reading Tutor's turn to pick a story, the tutor randomly selects a story at the student's current reading level that he has not yet read.

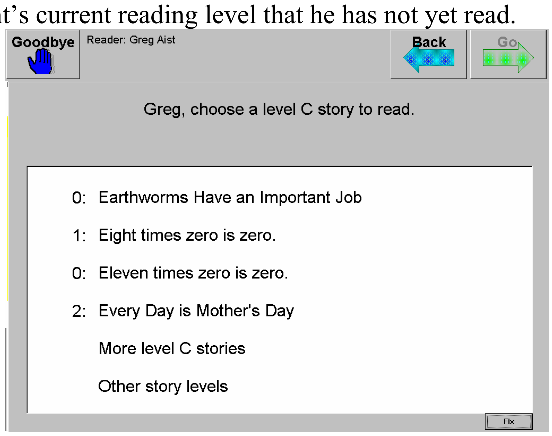

Figure 1. Student story choice menu in the Reading Tutor.

Once the student selects a story to read, the tutor presents it on the screen one sentence (or, for very long sentences, sentence fragment) at a time. Students read aloud to the tutor and are able to ask for help for challenging words. If the student does not like the story he is reading, he can return to the story choice menu to get a different story.

One method of determining whether the Reading Tutor's learner control is beneficial is to run a controlled study where one group of students always selects which story to read and the other group always has their story selected by the tutor. There are a few problems with this methodology. First, our *a priori* hypothesis is that there will be a relatively small effect on learning from students being allowed to select their own story to read. Therefore, a study would have to involve many users, be of considerable duration, or some combination of both, and would consequently be expensive to run.

Second, who would be the consumer for the results of such a study? The effect of learner control depends both on the types of learners and on the type of control. Learners with weaker metacognitive skills will (presumably) show less benefit from control than metacognitively advanced learners. Furthermore, perhaps some types of control are more effective than others. Would the results of such a study be informative for high school students studying geometry or for college students learning physics? Even if the learners were similar, would the results transfer across different types of learner control and to different domains?

Given the limited utility to other researchers, the option of running an expensive study is not an appealing one. Rather than trying paying a lot to get a definitive answer to a question that few others care about (how effective is the *Reading Tutor's* learner control?), instead we will concentrate on creating an easy recipe to enable researchers to answer the question of how effective is *their tutor's* learner control. If any researchers who were curious could cheaply test whether their system's learner control was beneficial, then enough such results might give us a better understanding of what types of control—and for what populations of students—there is a benefit. Even if the specific results for the effects of learner control in the Reading Tutor were not of interest to others, our method of obtaining the answer would be.

2.1 Learning Decomposition

To measure the impact of learner control, we use *learning decomposition* [8]. The idea of learning decomposition is to modify the learning curve equation to account for the impact of different types of learning opportunities. As can be seen in the standard form of the exponential learning curve in Equation 1a, the free parameter A represents how well students perform on their first trial performing the skill; e is the numerical constant (2.718); the free parameter b represents how quickly students learn the skill, and t is the number of practice trials the learner has had at this skill. This model can be solved with any non-linear regression package (we use SPSS 11.0). For this work, we are tracking student knowledge at the level of individual words, and therefore trials (t) are every time the student encounters the word in a story.

Traditionally, learning curves simply count up the number of prior learning opportunities. Learning decomposition instead separates learning opportunities into different types, as specified by the researcher. Equation 1b shows the form of the learning decomposition model. A, e, and b have the same meanings as in Equation 1a. Rather than simply counting all prior opportunities (t) equally, we split prior exposures into those that occurred in stories chosen by the Reading Tutor (t_1) and those chosen by the student (t_2). Equation 1b shows this decomposition. Since either the student or the tutor chose every story, it is the case that $t = t_1 + t_2$.

The additional free parameter, B, represents the relative impact of tutor control. If, for example, the best fitting parameter estimate for B is 0.5, then the development of student proficiency is best described by a model that weights learning opportunities in tutor selected stories as being worth half as much as opportunities in tutor selected stories. I.e. students learn half as quickly when the tutor selects the story they read. Note that our choice to allow the value of tutor selected stories to vary has no impact on the analysis. If we had instead estimated the effect of student selected stories by modeling $t_1 + B{*}t_2$ then we would have found that B=2.0 (i.e. students learn twice as quickly when they select the story—an identical finding). For our analysis, if $B>1$ then tutor choice results in additional learning over student control. If $B=1$, then there is no difference in learning. If $B<1$ then students learn more in stories that they pick.

Equation 1b are abridged for clarity, as we also consider the length of the word and how many prior help requests the student has made on the word. We will not discuss those aspects of the model in this paper.

$$performance = A * e^{-b*t} \qquad performance = A * e^{-b*(B*t_1 + t_2)}$$

Equation 1. (a) Exponential model, **(b)** Learning decomposition model of practice.

2.2 Measure of Reading Growth

We now instantiate what we mean by *performance* in Equation 1. When students read a word in the Reading Tutor, we record how long they took to read the word, whether the automated speech recognizer (ASR) thought the word was read correctly, and whether they asked for help on the word. We would like a performance measure to improve smoothly over time. Therefore, reading time is a better measure than accuracy or help request, both of which are binary outcomes and thus incapable of gradual improvement on individual trials. However, reading time is undefined for cases where the student did not attempt to read the word, and is artificially lowered for cases where the student first asks for help since the student can simply mimic the tutor's help. To avoid these confounds, we set reading time to be 3.0 seconds for cases where the student skips the word or asks for help. The time of 3.0 seconds is fairly high, since only 0.1% of readings exceeded that amount. Furthermore, reading times that exceeded 3.0 seconds were reduced to be 3.0 seconds. Although speech recognition is far from perfect and individual estimates of word reading time are inaccurate, errors should be randomly distributed. I.e. we do not expect either type of trial to benefit unfairly from the scoring inaccuracies and the results are an unbiased estimate of learner control.

We would also like our performance metric to reflect changes in student knowledge rather than temporary scaffolding effects. For example, if the student reads a story that has the word "divided" in 5 consecutive sentences (such stories exist in the Reading Tutor), by the fifth sentence one suspects the student is no longer reading the word in the normal sense, but is simply retrieving it from short term memory. Therefore, we only count the student's first exposure to a word in a day for the dependent variable. However, we count all of the exposures as possible learning opportunities (i.e. for the independent variable). So, in the above example, only the first sentence with "divided" would count as a training opportunity for the model, but all of the sentences would be counted to determine the correct values for t_1 and t_2. Fitting a simple exponential model that uses number of prior exposures to predict mean word reading time for a word with that many exposures had an R^2 of 76%. So our data are well described by an exponential model.

Our approach was to fit the nonlinear regression model in Equation 1b to each individual student. That is, we estimated the A, b, and B parameters for every student using only his data, and only considering students who read at least 20 words in the Reading Tutor. We used every word read by the student as training data for the model. We accounted for varying length of words by including a simple linear term to account for word length in characters. We expect considerable error in the parameter estimates of B, and so cannot make claims about the efficacy of learner control for individual students, but can use the noisy estimates to detect overall patterns across students. Our resulting data set had 346 students who had a total of 959,455 observations of learning (that is, first attempts for a student at reading a word that day). Overall these students read approximately 6.9 million words over the course of the year.

3 Results

To determine the effects of tutor- vs. student-selected stories, we compute the median B parameter estimates for all 346 students. The median value of B was 0.80, i.e. students only learned 80% as much in stories chosen by the tutor (alternately, 25% more in stories selected by the student). Note that this measure of increased learning does not consider short-term motivational effects. First, if a student read a word in a story that he selected, we only looked at encounters of the word on a later day as possible evidence of learning. So improved performance in the student selected story cannot be the cause of these results. Second, this study took place over an academic year (from September 2003 through May 2004), so the results are not a simple novelty effect.

Another possibility is that the results are not due to higher student motivation or engagement, but instead are due to students choosing stories that were more appropriate pedagogically. However, such is not the case. When it was the Reading Tutor's turn to pick a story, it selected a story at the student's estimated reading level that he had not previously read. Students could select stories at any level of difficulty and could reread stories. Students and the Reading Tutor selected stories at nearly the same level of difficulty. For the 93347 student-chosen stories in which they spent at least one minute, the mean difficulty was 2.27 (where 2 is second grade reading level and 3 is third grade reading level). For the 100417 tutor-chosen stories in which the student spent at least one minute, the mean difficulty was 2.01. So the difficulty levels are comparable across selection types. Furthermore, the student ability to reread stories likely hurt their learning, since we have found that rereading previously read stories leads to less efficient learning [8].

Given that students learn more when they select which story to read, an obvious question is whether the benefit outweighs the cost. As a simple model, we consider how much time students spent reading stories (presumably productive learning time) vs. how much time they spent picking stories (presumably non-productive learning time—at least for learning how to decode words). In this data set, students read for 3053 hours and spent 503 hours selecting which story to read. If the tutor instead chose a story and forced the student to read it, those 503 hours could have instead been spent reading, resulting in 3053 + 503 = 3556 hours of instruction. Unfortunately, since stories selected by the tutor only result in 80% as much learning, the amount of comparable time on task would be 3556 * 0.8 = 2845 hours—fewer than the 3053 hours that students actually spent reading. So it appears that allowing student to select activities is worth the time cost. Actually, this analysis is an underestimate of the benefit since there are probably benefits to persistence (how long students are willing to use the tutor) that are unaccounted for. Also, as anyone who has observed actual field conditions for computer tutors can attest, the part about forcing students to read a story is problematic, at best, to instantiate.

3.1 Median as a Measure of Central Tendency

There was considerable variance as the per-student estimates of B ranged from -201 to 2185. Since some students have relatively few observations, their estimated value of B is rather inaccurate. One approach would be to trim or cap outliers and then take the mean. However, this approach requires first determining what constitutes an outlier. Presumably the student who the model claims learned 2000 times as much is a result of a poorly estimated parameter value and should be ignored, but what of the students who learned 5 times as much as a result of tutor control? 3 times as much? Triple the rate of learning is implausible, but possible. Rather than create an arbitrary threshold for when a student is an outlier, we simply use the median as a measure of central tendency. This process compresses the effects of outliers. For example, the median would be the same if the student whose B parameter was 2000 was instead 2. In contrast, the mean is sensitive to magnitude, and in fact the mean value of *B* is an implausible 20.2.

There is a second difficulty with using the mean. Imagine a population consisting of three students. One student learns three times as much when the tutor picks the story (*B*=3) while the two other students learns one-third as much when the tutor picks the story (*B*=0.33). The mean of this sample is (3 + 0.33 + 0.33) / 3 = 1.22—a counter intuitive result given that 3 and one-third are exactly opposite effects, and twice as many students are having negative results. Therefore, we feel the use of median as a measure of central tendency is appropriate in this context.

3.2 Which Students Benefit from Learner Control?

We have established that students in the Reading Tutor benefit from a degree of learner control, and that this control is worth the time it takes away from learning. The final question we address is whether all students benefit from learner control? If certain students do not benefit from or need the motivational effect of learning control, we can make their learning more effective by having the tutor select materials for them.

We present two approaches for finding such students. The first approach is top-down and disaggregates the results by the student's grade level. We found that the 130 first grade students (typically 6 years of age) had a median *B* of 0.98, the 134 second graders a 0.89, and the 70 third graders a 0.49. We had only 8 and 4 students in grades four and five, respectively, so the medians were not reliable. This trend suggests that younger readers are insensitive to the material they are reading. As readers become more advanced, they develop preferences in what they wish to read and they are less engaged in tutor selected stories.

The second approach is bottom-up, and instead treats students as individuals and attempts to discover similarities between students who benefit from learner control vs. those who do not. Our approach is to use the student's estimated *B* score as a noisy category label and treat the task as one of classification. If a student's *B* score was below 0.9, this student was labeled as benefiting from learner control (190 students). If the student's *B* score was above 1.1, the student was labeled as benefiting from tutor control (133 students). The remaining 23 students were not used for the training. For our classifier, we used logistic regression with the following independent variables:

- The student's grade (covariate)
- The student's grade relative Woodcock Reading Mastery Test composite score [9] (covariate)
- The student's gender (factor)
- Whether the student was receiving learning support services (factor) which took on three levels: yes (32 students), no (169 students), and unknown (145 students)

We chose those dependent variables since they are likely to be known by teachers and researchers at the beginning of the school year. Only the second variable, the grade relative Woodcock Reading Mastery Test score, is difficult to obtain, and we suspect that most standardized test scores for the domain would do just as well.

The model's Nagelkerke R^2 was a rather poor 0.045, with the only statistically reliable independent being the student's gender. Apparently girls were more likely to benefit from tutor control, while boys are more likely to benefit from learner control.

One possible explanation is that students tend to select stories that are less than optimal from a learning perspective. If the Reading Tutor is selecting pedagogically better stories than the student, then students who are engaged regardless of the material being read will benefit more from tutor control. Perhaps young girls are more interested in reading than young boys? In fact, we found that girls scored reliably higher ($P<=0.001$, two-tailed t-test) on measures of both academic and recreational reading interest [10].

One explanation for the low model fit is the noisy class labels. The per-student estimates of *B* are individually suspect, and we are relying on the fact that we have hundreds of estimates to try to derive a pattern. That the model is doing a better job at separating students than the fit statistics suggest is evident from looking at its predictions. Students the model predicted would benefit from learner control (N=261) had a median *B* parameter of 0.66, students the model predicted would be hurt by learner control (N=61) had a median *B* of 1.23—a wide degree of separation.

4 Contributions, Future Work, and Conclusions

This paper's main contributions is measuring the effect that perceived learner control in an ITS has on student learning. Prior work, such as [6] examined the effects of learner customization and found that fairly minor modifications (icons used to represent the student, referring to friends of the student in background text, etc.) resulted in statistically reliable gains in learning. However, this study was of short duration: only four experimental sessions. It is unclear whether cosmetic modifications will succeed in maintaining student interest over the course of an entire academic year. Our study shows that a fairly minor choice of which story to read next results in noticeable differences in the student learning of the words in the story. Our study also determines for whom such a benefit is more likely to occur. Namely, that boys are more likely to benefit from the Reading Tutor's perceived control than girls.

This work also introduces a methodology, learning decomposition, for investigating issues involving the student's affective state. This approach required no additional data collection. In fact, this analysis was not conceived of until roughly two years after the study ended in Spring 2004. The approach of learning decomposition is broadly applicable, and it is not simply appropriate for studying affective factors such as engagement or motivation. Our prior work with this technique [8] investigated the impact of cognitive factors (spaced practice) and suspected reading curriculum issues (rereading the same story vs. reading new stories).

Although this work makes a good start in creating a framework for investigating effects of learner control, it is only a start. First, this analysis has not been replicated via a classic controlled study. Our prior belief was that the difference in learning gains would be small, and a controlled study unwarranted. A 25% increase in learning is substantial, and worth exploring further. Second, our analysis only pertains to a specific instantiation of selecting the next task. Would constraining the space of learner decisions still result in the same benefit? For example, imagine if we presented many fewer stories for the student to choose among. This design change would reduce the time it takes students to select stories, resulting in more time on task. But would we get the same increase in learning as we found in this study?

One other limiting factor is scope. We have analyzed one type of learner decision, what story (i.e. problem) to work on next. Do other types of learner control in ITS show similar benefits? How much learner control is needed? Is one choice at the beginning of the session with the tutor sufficient, or do students require periodic choice points? There is a broad scope of possible instantiations of learner control, of which this paper examines only one. One final limitation is that learning decomposition requires fine-grained markers of student progress such as performance data on individual problems.

There is also a limitation that not all students are engaged and on-task when the tutor selects stories. For example, students finished 47% of stories that they selected vs. only 38% of stories selected by the tutor—however, we used all of the words the student read, even those in uncompleted stories, to train our model. Presumably this differential attrition was from students who were not motivated by the tutor's choice. Therefore, the estimate of learning in tutor-chosen stories is from a group of students who were generally intrinsically more motivated to read than in student-chosen stories. We do not see a good way to address this issue. However, this bias causes tutor-chosen stories to look more effective for learning than they actually are, which would weaken the main result in the paper that student-chosen stories result in more learning. Thus the real difference is likely stronger than what we report.

In conclusion, we have shown that student choice—at least for our instantiation of it in the Reading Tutor—not only increases student motivation, it also increases student learning for activities the student chose to perform vs. those he did not select. We have also found that older children and boys are more sensitive to learner control.

Acknowledgements. This work was supported by the National Science Foundation, 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

References

1. 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.
2. 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. p. 195-198 Venice, Italy.
3. 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. p. 292-303 Montreal: Berlin: Springer Verlag.
4. Beck, J. *Engagement tracing: using response times to model student disengagement*. in *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED 2005)*. 2005. p. 88-95 Amsterdam: IOS Press.
5. Becker, D.A.A. and M.M. Dwyer, *Using hypermedia to provide learner control.* Journal of Educational Multimedia and Hypermedia, 1994. **3**(2): p. 155-172.
6. Cordova, D.I. and M.R. Lepper, *Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice.* Journal of Educational Psychology, 1996. **88**(4): p. 715-730.
7. Aist, G. and J. Mostow, *Faster, better task choice in a reading tutor that listens*, in *The Path of Speech Technologies in CALL: Foundations, Prototypes, and Evaluation*, V.M. Holland and F.N. Fisher, Editors. in press, Routledge Publishers: New York.
8. Beck, J. *Using learning decomposition to analyze student fluency development*. in *ITS2006 Educational Data Mining Workshop*. 2006. p. Jhongli, Taiwan.
9. Woodcock, R.W., *Woodcock Reading Mastery Tests - Revised (WRMT-R/NU)*. 1998, Circle Pines, Minnesota: American Guidance Service.
10. McKenna, M.C., D.J. Kear, and R.A. Ellsworth, *Children's attitudes toward reading: a national survey.* Reading Research Quarterly, 1995. **30**: p. 934-956.