The Effects of Adaptive Sequencing Algorithms on Player Engagement within an Online Game

Derek Lomas[1], John Stamper[1], Ryan Muller[1], Kishan Patel[2], and Kenneth Koedinger[1]

[1] Carnegie Mellon University
{dereklomas,jstamper,rmuller,koedinger}@cs.cmu.edu
[2] Dhirubhai Ambani Institute of Information and Communication, Gujarat, India
kishan_patel@daiict.ac.in

Abstract. Using the online educational game *Battleship Numberline*, we have collected over 8 million number line estimates from hundreds of thousands of players. Using random assignment, we evaluate the effects of various adaptive sequencing algorithms on player engagement and learning.

Keywords: games, number sense, engagement, adaptive sequencing.

1 Introduction

Number line estimation accuracy is highly correlated with math achievement scores in grades K-8 (Siegler, Thompson, Schneider, 2011). To promote practice with number line estimation, we have developed *Battleship Numberline,* a game involving estimating the location of ships on a number line. Using this game, we have collected over 8 million number line estimates from several hundred thousand online players. The order of instructional items in the game is typically presented at random, but we hypothesize that an adaptive sequence will result in an improved learning experience. Adaptive instructional sequences are best known for increasing the efficiency of learning [2]. However, Pavlik et al. [3] reported that students tended to chose an adaptive sequence of foreign language instructional items over a random sequence of items. We further explore this phenomenon by investigating whether adaptive sequences can increase motivation to engage in a learning activity.

2 Adaptive Sequences

Conati et al. [1] describe using Bayesian Knowledge Tracing (BKT) to promote learning in an educational game. However, many games use far simpler algorithms to promote learning and player interest; for instance, they may require a player to perform flawlessly on a level before progressing to the next. Could simpler adaptive algorithms achieve comparable performance to Bayesian Knowledge Tracing? Specifically, could they produce comparable learning (pre-post test gain) and player engagement (duration of intrinsically-motivated play)?

In our implementation of BKT, we developed a knowledge component model with five knowledge components (KC). The parameters for the model were developed

S.A. Cerri and B. Clancey (Eds.): ITS 2012, LNCS 7315, pp. 591–593, 2012.

based on data collected from a prior classroom study involving 150 students in 4th-6th grade. These parameters included the probability of existing knowledge (L0), learning rates (T), and the probability for slipping (S) and guessing (G). The sequencing algorithm worked by randomly choosing an item belonging to the KC with the highest probability of being known, so long as it was below the threshold of .9 probability of being known. When a KC exceeded .9, it was removed from the sequence. Once all KCs in the level exceeded .9, the level was over.

The Difficulty Ladder (dLadder) is an adaptive sequencing algorithm that requires mastery of easier items before allowing progress to more difficult items. Based on the same dataset from which the BKT parameters were derived, the items in the instructional sequence were divided into 5 bins of difficulty, each with 4 items. Players began in the easiest bin; if they were correct twice in a row, they advanced to the next more difficult bin. If they were incorrect twice in a row, they went back to the previous, less difficult bin. When the player completed the hardest bin, the level was over. A high performing player could complete the ladder in only 10 trials.

Naïve ITS is based on the idea that a successful response tends to generate more learning than an unsuccessful response. To promote success, if a player gets an item incorrect, they are given another opportunity to attempt the item after a delay of one other item. The delay of one trial facilitates working memory retrieval without making the task trivially easy (as it might be if there was no delay). Once the player gets every item correct at least once, the level is over.

The random sequence randomly presents (without replacement) one of 20 different fractions. Unlike the adaptive sequences, the random sequence is not affected by the player's prior performance.

3 Experiment 1: Structure, Participants and Metrics

The adaptive sequencing experiment involved randomly assigning 1087 players to one of sixteen different level sequences representing four different experimental conditions (BKT, Difficulty Ladder, Naïve ITS, & Random) with four different pre/posttest form combinations (A-B, B-C, C-D, D-A). Each level sequence consisted of a pretest level, a level with one of four sequencing algorithms, a post-test level, and then additional levels of the same sequencing algorithm (so that patterns of extended play could be compared over the different algorithms). The pre/post tests involve four fraction estimation problems, presented fully within the context of the game.

Our participants are anonymous online players who freely access our game through the educational portal Brainpop.com. Despite this anonymity, we can infer from the demographics of Brainpop.com that our users are likely to be third to eighth grade students, probably playing in a classroom setting. Brainpop.com offers a number of different educational games. We assume that students are free to stop playing *Battleship Numberline* at any time; indeed, over 50% of students play less than 10 trials.

In this study, we define engagement as the number of trials that a player chooses to play, as this is believed to reflect the players intrinsic motivation to participate in the gameplay sequence. We measure learning as the gain from pretest to posttest.

Table 1. Initial conditions of experiment

	Completed Pretest	Pretest Av.	Av. # of Trials	Median # of Trials	% playing > 40 Trials
BKT	265	23% (.42)	25(30)	14	20%
DLadder	267	23% (.42)	29(35)	16	25%
NaiveITS	279	26% (.44)	30(37)	16	24%
Random	276	23% (.42)	24(22)	15	21%

Table 2. Here, data is presented only for the players that completed the posttest. Gain is significant from pre to post test over all conditions (p<.02, p<.01, p<.001) using a paired t-test.

	Completed Posttest	Pretest Av.	Posttest Av.	Median # of Trials
BKT	0	n/a	n/a	n/a
DLadder	22	46% (.50)	65%(.48)	30.5
NaiveITS	55	31% (.46)	47%(.50)	49
Random	103	25% (.43)	37%(.48)	28

4 Discussion

The data presented here suggests a modest effect from the sequencing algorithms. Unfortunately, learning gains are impossible to compare directly, without statistically correcting for the substantial rates of attrition. Our BKT algorithm apparently set the bar too high—no players in this sample actually completed the level, despite some players completing more than 100 trials. Future work will involve tuning the parameters of the BKT algorithm, developing more comparable measures of learning, and validating our online engagement construct in a classroom setting.

References

1. Conati, C., Zhao, X.: Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, pp. 6–13. ACM (2004)
2. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modelling and User-Adapted Interaction 4(4), 253–278 (1995)
3. Pavlik Jr., P., Bolster, T., Wu, S.-M., Koedinger, K., MacWhinney, B.: Using Optimally Selected Drill Practice to Train Basic Facts. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 593–602. Springer, Heidelberg (2008)
4. Siegler, R.S., Thompson, C.A., Schneider, M.: An integrated theory of whole number and fractions development. Cognitive Psychology 62(4), 273–296 (2011)