

Does the Length of Time Off-Task Matter?

Daniel A Roberge
Anthony C Rojas

Interactive Qualifying Project Report completed in partial fulfillment
of the Bachelor of Science degree at
Worcester Polytechnic Institute, Worcester, MA

Submitted to:
Professor Ryan S. J. d. Baker (advisor)

January 24, 2012

Advisor Signature

Abstract

We investigate the relationship between a student's time off-task and the amount that he or she learns to see whether or not the relationship between time off-task and learning is a more complex model than the traditional linear model typically studied. The data collected is based off of students' interactions with Cognitive Tutor learning software. Analysis suggested that more complex functions did not fit the data significantly better than a linear function. In addition, there was no evidence that the length of a specific pause matters for predicting learning outcomes; e.g. students who make many short pauses do not appear to learn more or less than students who make a smaller number of long pauses. As such, previous theoretical accounts arguing that off-task behavior primarily reduces learning by reducing the amount of time spent learning remain congruent with the current evidence.

Acknowledgements

We would like to thank Art Graesser for helpful comments and suggestions, Angela Wagner for her help collecting the data re-analyzed here, and Professor Ryan Baker for all his helpful ideas and revisions that aided in the completion of this paper.

Table of Contents

Abstract	2
Acknowledgements	3
Introduction.....	5
Data	6
Analysis.....	8
Percent of Time Off-Task.....	8
Number of Brief/Lengthy Times Off-Task.....	10
Discussion and Conclusion	12
References.....	14

Introduction

Today, students are interacting with technology of various forms more than ever during learning. One context where this is occurring is in middle school and high school mathematics classes, where learning increasingly occurs from students using educational software in a classroom with a teacher present. One such form of educational software which is used progressively more in the United States is Cognitive Tutor software [11], where student learning is individualized based on assessment of the student's current learning and the factors leading to a specific student error.

Cognitive Tutors are now used in more than 6% of U.S. secondary schools. Cognitive Tutors have been shown to enable individual students to learn at their own pace, while empowering teachers to spend instructional time in one-on-one teaching episodes with the students who are struggling most [16]. Educational software like Cognitive Tutors provide extensive logs of student performance [12], enabling not only more effective learning for individual students, but supporting analysis of student learning over time, using methods from learning analytics [17] or educational data mining [3].

In this paper, we study the relationship between a student's learning gain and his or her off-task behavior. One key model for this relationship is Carroll's Time-On-Task hypothesis. This hypothesis argues that off-task behavior reduces learning by reducing the amount of time spent on-task [7]. However, there are several factors that may complicate this relationship. In particular, it is possible that taking a short break may improve cognition afterwards [cf. 13]. Hence, short pauses may impact learning differently than longer pauses. In addition, it is possible that a qualitative difference may be seen between students who go off-task for a break, and students who are more fundamentally disengaged; hence, students who are off-task for large proportions of time may have greater reduction of learning than could be anticipated through a simple linear model. For this reason, we investigate the differences between several models of the relationship between off-task behavior and learning,

leveraging both quantitative field observations of off-task behavior [cf. 2] and an automated detector of off-task behavior [cf. 4] to measure both overall prevalence of off-task behavior and the duration of individual episodes. We looked at the percent of time off-task and the number of brief and lengthy off-task episodes to study the relationship between these factors and learning gain.

Through the use of statistical analyses on medium-sized educational data sets, a form of learning analytics [cf. 17], we can better understand how off-task behavior influences learning and under what conditions off-task behavior influences learning differently. A learning analytics analysis using similar methods includes research on student activities during writing [cf. 5].

Past studies conducted on students using educational software have generally shown a negative correlation between off-task behavior and learning during the use of the software [8, 9, 15]. A similar pattern has been seen when studying these relationships outside of technology, generally finding a negative relationship between learning and off-task behavior (see [6] for an extensive review of this literature). However, these studies have typically not explored non-linear relationships. A study done by Karweit and Slavin, however, found that changing the length of observation periods affected the strength of the relationships between off-task actions and learning [10]. This supports the notion that the length of off-task episodes may be predictive of student learning, as well as the overall quantity of time spent off-task.

Data

The students in this experiment used educational software in the domain of scatterplots, a subject taught in the data analysis portion of middle school mathematics in the United States. Initially, students took a pre-test to determine how well they knew the material at hand. Afterwards, the students interacted with a Cognitive Tutor lesson teaching this topic [1], for approximately 80 minutes

apiece. Finally, a post-test assessment was given to evaluate the students' progress. Full details on the assessments are given in [1]. 186 students completed the pre-test, the tutor activity, and the post-test. These students were drawn from multiple previous studies in separate years [cf. 8], but each used the same tutor software under the same conditions (in some studies, these students served as the control condition which was compared to a modified version of the tutor – students using modified versions of the tutor are not analyzed in this paper).

Data on student off-task behavior was gathered using two methods. While students engaged in the cognitive tutor classroom [11], two observers recorded student's behaviors using quantitative field observations [2]. The students were observed using peripheral vision in order to decrease potential observer effects, in a sequence of 20-second long observations cycled across students. In each observation, the student's behavior was noted in terms of whether it involved off-task behavior [2, 4], in order to compute the percentage of time each student was off-task. Off-task behavior was defined as any of the following: off-task conversation (talking about anything other than the subject material), off-task solitary behavior (any behavior that did not involve the tutoring software or another individual, such as reading a magazine or surfing the web), and inactivity (such as staring into space, or the student putting his/her head down on the desk, for at least 20 seconds – brief reflective pauses by a student actively using the software were not counted as off-task). Other behaviors such as actively working in the software, collaborating with other students, and gaming the system (intentionally misusing the software in order to successfully complete problems [cf. 2]) were not counted as being off-task.

The second method used an automated detector of off-task behavior developed using data mining [4], built using the field observations as ground truth. The off-task detector is a latent response model used to infer exactly when off-task behavior occurs, from features of individual student actions

and recent student behavior before those actions. The detector was shown to achieve a correlation over 0.5 to the proportion of off-task behavior observed, under student-level cross-validation. As such, we use the detectors' inferences as components in further analysis, a process termed "discovery with models" [3]. As off-task episodes can sometimes be caught by the detector from the behavior occurring shortly after the actual off-task episode (in which case they are identified by very quick actions coming after the a long pause) [4], we label each off-task pause with the length of the longest pause in the sequence of 5 student actions considered by the detector when making an inference.

When looking at the predicted models for the post-test score, there were two options. The traditional, linear equation was given by:

$$Post = 0.737 OT + 0.716$$

The sum of squared residuals (SSR) produced by this model was 27.169. Since the data was generally nonlinear, it seemed appropriate to look at the predicted post-test score from a quadratic standpoint. This model (given below) had an SSR of 26.895.

$$Post = -0.75 OT^2 + 0.05 OT + 0.68$$

These SSR's provide some insight into the nonlinear nature of the dataset, which is explored in greater depth below.

Analysis

Percent of Time Off-Task

The first relationship that we considered was the percentage of time the student spent off-task while using the cognitive tutor. Post-test score was the indicator used as an assessment of the student's eventual learning.

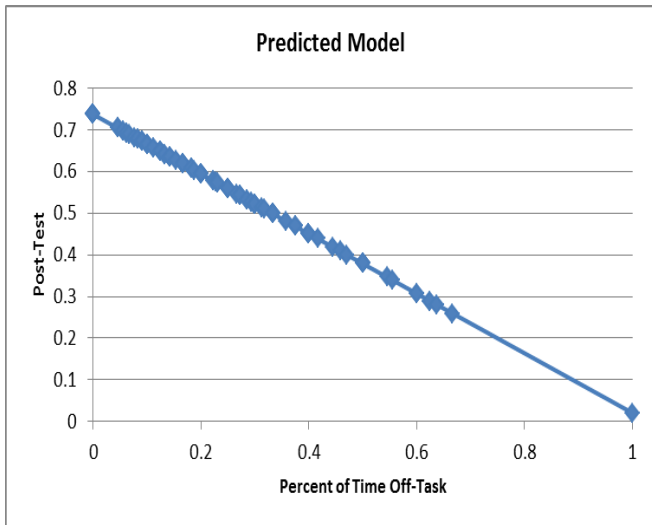


Figure 1: The predicted post-test score (from the linear model below) compared to the percent of time-off-task.

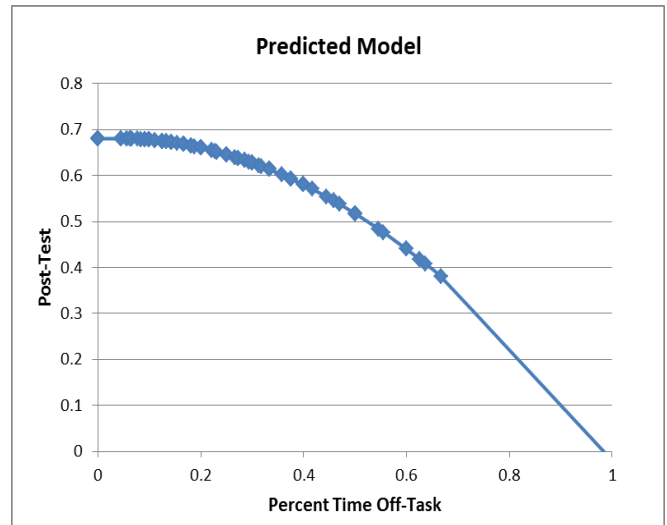


Figure 2: The predicted post-test score (from the non-linear model below) compared to the percent of time-off-task.

With the analyses in this section, we measure each student’s proportion of off-task behavior, using the field observations, as this is the most standard method for assessing off-task behavior, used by researchers for decades [e.g. 2, 4, 6, 9, 10]. A student’s proportion of off-task behavior is statistically significantly negatively correlated to their post-test score, $r = -0.229$, $F(1,184) = 10.150$, $p < 0.01$. The student’s pre-test score was statistically significantly positively correlated to the post-test as well, $r = 0.299$, $F(1, 184) = 18.007$, $p < 0.001$.

In order to study the relationship between off-task behavior and learning, we can analyze the relationship between off-task behavior and the post-test, while controlling for the pre-test. The best-fitting linear model of this relationship is:

$$Post = 0.273Pre - 0.394OT + 0.617$$

Within this model, the off-task term was significantly different than chance, $t(185) = -2.412$, $p = 0.017$.

We can also investigate a non-linear model (shown in Figure 2), including the percentage of time off-task, squared, which produced the best-fitting equation:

$$Post = 0.275Pre - 0.848OT^2 + 0.598$$

For the off-task squared term, $t(185) = -2.295$ and $p = 0.023$.

Hence, both linear and quadratic models based on off-task behavior are significant predictors of student off-task behavior. In order to investigate whether one model is significantly better than the other, we compare the models using the Bayesian Information Criterion for Linear Regression, BIC' [14]. BIC' is a formula used to see how well a specific model predicts the data given the number of parameters (e.g. models with more parameters should achieve better fit simply by chance). It can also be used to compare two non-nested models of the same dependent measure, so we can use it to compare the models predicting post-test using the proportion of off-task behavior in a linear fashion, and the proportion of off-task behavior, squared. In the first case, where pre-test and percent of time off-task predict post-test, the BIC' produced a result of -12.685; while pre-test and percent of time off-task squared produced a result of -12.255. Although the regression model that uses time off-task squared produces a higher r value, the difference in the BIC' of the two models is only 0.430. This indicates that the two models are not statistically different, which would be indicated by a difference of 6 or greater [14]. Hence, there is not a strong justification for preferring a non-linear model of the relationship between off-task behavior and learning, to a linear model, although there is a trend in that direction.

Number of Brief/Lengthy Times Off-Task

A second question is whether lengthy pauses impact learning in a different fashion than brief pauses. It has been shown that breaks in the workplace can reduce mental fatigue and ultimately lead to better employee performance on cognitive tasks [13]. Thus, it is possible that the number of times a student has either brief or long pauses affects their learning gain.

Within the analyses in this section, we use the automated detector of off-task behavior rather than the field observations of off-task behavior. The type of field observation used when the data was

first collected – round-robin observations of an entire classroom by one or more field observers – gives a useful representation of the total proportion of time each student is off-task, but it does not shed light on how long individual episodes of off-task behavior are. By labeling every student action across the entire session (and gaps between actions) as to whether it is off-task or not, automated detectors allow us to analyze the length of individual episodes of off-task behavior.

An analysis predicting learning using the total number of times off-task gave the following results. These two variables had a correlation of 0.053, $F(1,184)=0.519$ and $p=0.472$. When pre-test was included as a covariate, there was not a substantial difference for the term indicating the number of times off-task: $t(185)= 0.027$, $p= 0.979$. Hence, these results show that the total number of off-task episodes as indicated by the detector is not predictive of learning; however, when broken into brief and lengthy episodes there may be a significant relationship.

To investigate the difference between lengthy and brief pauses, we split the off-task episodes, as assessed by the detector, by their length in two fashions. First, a median split was conducted in terms of the length of an off-task episode. Episodes that were shorter than the median 65.9 seconds were classified as “brief” off-task episodes, whereas episodes that were longer than the median were classified as “lengthy” off-task episodes. We also conducted a quartile split, and compared the shortest-time quartile (less than 26.0 seconds) of the off-task episodes to the longest-time quartile (longer than 124.9 seconds) of the off-task episodes. In this manner, we can examine the difference in the correlation between learning gain and the number of times a student spent off-task, between brief off-task episodes and lengthy off-task episodes.

We first analyze the median split. The relationship between the number of off-task behavior episodes shorter than 65.9 seconds and the post-test was not significant, $r= -0.028$, $F(1,184)=0.142$, $p=0.706$. The relationship between the number of off-task behavior episodes longer than 65.9 seconds

and the post-test was surprisingly also not significant, $r = -0.056$, $F(1,184)=0.586$, $p=0.445$. These patterns remained non-significant even when included the pre-test as a covariate. There was no difference between these two models, $BIC' = 0$.

A similar pattern is seen when comparing the top quartile and bottom quartile. The relationship between the number of off-task behavior episodes shorter than 26.0 seconds and the post-test was not significant, $r = -0.101$, $F(1,184)=1.897$, $p=0.170$. The relationship between the number of off-task behavior episodes longer than 124.9 seconds and the post-test was surprisingly also not significant, $r = -0.077$, $F(1,184)=1.098$, $p=0.296$. This pattern remained non-significant even when pre-test was included as a covariate. The difference between these two models was not significant, $BIC' = -1.501$.

Discussion and Conclusion

In general, this paper replicated previous findings showing a negative relationship between off-task behaviors assessed using field observations and learning. Our results showed that there was a significant relationship between the percentage of time the student spent off-task and learning. There was also evidence for a relationship between the proportion of off-task behavior squared, and learning. However, the difference between the quadratic and linear models of this relationship was not significant; suggesting that more complex models than the model hypothesized by Carroll may not be justified. There was also not strong evidence that brief off-task episodes impact learning differently than longer off-task episodes.

One surprise in the findings was that the number of episodes identified by the off-task detector was not predictive of learning, either for brief episodes or lengthy episodes. Previous analyses have found a significant relationship between the proportion of off-task behavior identified by the detector and learning [e.g. 3]. Those analyses were conducted in a broader data set, including data from other

versions of the same learning software, and other tutor lessons. In this paper, we analyzed a more focused data set, in order to explore different models in detail without needing to consider this type of factor. But it is possible that features specific to this sub-set of the data or the associated tutor lesson led to the null result seen here. Therefore, it may be valuable to replicate these analyses in a larger data set.

In summary, the data in this study appeared to accord with Carroll's time on-task hypothesis [7]. Currently, there is not sufficient evidence to suggest that a more complex relationship exists between learning gain and off-task behavior, in terms of the temporal aspects of off-task pauses. However, this issue may be worth further investigation in additional data, before this result can be considered conclusive.

References

- [1] Baker, R.S., Corbett, A.T., and Koedinger, K.R., Learning to Distinguish Between Representations of Data: A Cognitive Tutor That Uses Contrasting Cases. in *International Conference of the Learning Sciences*, (2004), 58-65.
- [2] Baker, R.S., Corbett, A.T., Koedinger, K.R., and Wagner, A.Z., Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". in *ACM CHI 2004: Computer-Human Interaction*, (2004), 383-390.
- [3] Baker, R.S.J.d. and Yacef, K. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1 (1), 3-17.
- [4] Baker, R.S.J.d., Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. in *ACM CHI 2007: Computer-Human Interaction*, (2007), 1059-1068.
- [5] Blikstein, P., Using Learning Analytics to Assess Students' Behavior in Open-Ended Programming Tasks. in *Learning Analytics Knowledge Conference* (Banff, Canada, 2011).
- [6] Caldwell, J.H., Huitt, W.G., and Graeber, A.O. Time Spent in Learning: Implications from Research. *The Elementary School Journal*, 82 (5), 171-180.
- [7] Carroll, J. A Model for School Learning. *Teachers College Record*, 64, 723-733.
- [8] Cocea, M., Hershkovitz, A., and Baker, R.S.J.d., The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? in *14th International Conference on Artificial Intelligence in Education*, (2009), 507-514.
- [9] Gobel, Peter. Student Off-task Behavior and Motivation in the CALL Classroom. *International Journal of Pedagogies and Learning*, 4 (4), 4-18.

- [10] Karweit, N. and Slavin, R.E. Measurement and Modeling Choices in Studies of Time and Learning. *American Educational Research Journal*, 18 (2), 157-171.
- [11] Koedinger, K.R. and Corbett, A.T. Cognitive tutors: Technology Bringing Learning Science to the Classroom. in Sawyer, K. ed. *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press, 2006, 61–78.
- [12] Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., and Stamper, J. A Data Repository for the EDM Community: The PSLC DataShop. in Romero, C., Ventura, S., Pechenizkiy, M., and Baker, R.S.J.d. eds. *Handbook of Educational Data Mining*, CRC Press, Boca Raton, 2010, 43-56.
- [13] Meijman, T.F. (1995). Mental Fatigue and the Temporal Structuring of Working Times. in *Human Factors and Ergonomics Society Annual Meeting*, (1995), 789-793.
- [14] Raftery, A. Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111-163.
- [15] Rowe, J., McQuiggan, S., Robison, J., and Lester, J. Off-Task Behavior in Narrative-Centered Learning Environments. in *14th International Conference on Artificial Intelligence and Education*, (2009), 99–106.
- [16] Schofield, J.W. *Computers and Classroom Culture*. Cambridge University Press, New York, 1995.
- [17] Siemens, G. What are learning analytics?, August 25, 2010. Retrieved October 11, 2011, from Elearnspace: <http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>.