Predicting Off-task Behaviors in an Adaptive Vocabulary Learning System

SungJin Nam
School of Information, University of Michigan

INTRODUCTION

❑ ENGAGEMENT AND LEARNING

- Engagement is an important predictor of learning and is a multidimensional construct.
- **Capturing engagement in digital learning environments** is important for designing more effective and efficient learning experience.
- Model **a particular subset of student's off-task behaviors** in a vocabulary system use.

❑ KEY RESEARCH QUESTIONS

- Identifying important **predictive features of specific class of off-task behaviors** in a vocabulary learning system's log data.
- **Evaluating different modeling methods** that can more accurately predict off-task behaviors.
- **Suggesting effective adaptive strategies** for vocabulary learning systems that will help to sustain student's engagement and thus improve their learning outcomes and experience.

❑ AUTOMATED ASSESSMENT

- The current study uses **DSCoVAR** [2], **an intelligent tutoring system** for **vocabulary learning**.
- DSCoVAR includes automated assessment of performance, using free-response data from a meaning-generation task [3-4].

CURRENT STUDY

❑ PARTICIPANTS

- Native English speakers (**4th-6th grade,** 16 boys) from Falk Laboratory School [5,6].

❑ TASK

- Students provided **Familiarity Ratings** for 60 SAT level words (ratings: 53%, 23%, and 26%).
- Familiarity Ratings were followed by the **Meaning-Generation task** (Fig 2).

❑ OUTCOME MEASURE

- Two **human raters** labeled "**off-task"** responses (inter-rater agreement was *Kappa* = 0.695).
- **Instructions** (based on Baker et al. [1]):
  - "The response seems less serious or less relevant for a given target word"
  - "The response was part of repetitive responses over different question items"
  - "The response was part of repetitive false submissions"

❑ FEATURES

- **Real-time variables (RTV)** : Features obtained from a single response.
- **Context-based variables (CTV)** : Features from historical responses and other students.
- Using stepwise algorithm to select features.

ANALYSIS & RESULTS

Off-task Behavior Model (■= RTV / ■ = CTV / ■ = Random Variables)

RTV model (AUC: 0.918)

1 + (RT_Making + Resp.Length + Ort.Similarity)+ (1 | Target) + (1 | Subject)

RTV + CTV model (AUC:0.970)

1 + (Resp.Length + OrtographicSimilarity) + (SemanticSimilarity.p3 + OrthoSimilarity.p7) + (1 + Familiarity | Target) + (1 | Subject)

❑ FIXED EFFECT RESULTS

- **Real-time variables (RTVs)**
  - **RT_Making:** Shorter response time for typing in more likely to be an off-task behavior.
  - **RespLength:** Shorter responses were more likely to be an off-task behavior.
  - **OrthoOverlap:** Responses that were orthographically similar to the target word were more likely to labeled as an off-task behavior.
  - **SpellErr**: # of spelling errors.
  - **FormErr**: # of response format errors.
  - **RT_Start**:Time spent before initiating the response.
- **Context-based variables (CTVs)**
  - **SemanticDistance_prev.3**: Responses that were semantically similar with previous 3 responses were less likely to be an off-task behavior.
  - **OrthoRepetition_prev.7:** Responses that were orthographically similar with previous 7 responses were less likely to be an off-task behavior.
  - **pFlag_prev.X**: Proportion of off-task responses in previous.
  - **TargFlags.prev.X**: Average proportion of off-task responses for previous X trials from other students.
  - **TargetFlags**: Proportion of off-task responses for the target word from other students.

❑ RANDOM EFFECT RESULTS

- **Familiarity**: Words rated as **unknown** were more likely to elicit **off-task responses**.
- **Variability across items & students:** highlights the importance of models that capture multiple sources of variance, including random as well as fixed effects.

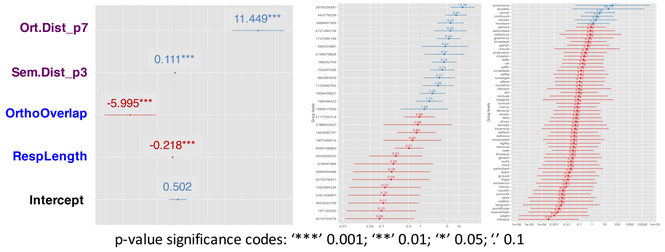


Figure 3. Coefficients for Fixed Effects (left) and Random Effect intercepts showing variation across students and target words (right; also see Fig. 1).

Table 3. Fixed Effects (GLMM: RTV+CTV).

Variables	Model 1 (RTV)			Model 2 (RTV + CTV)		
	Coeff	SE	z	Coeff	SE	z
(Intercept)	3.45**	1.21	2.84	0.50	0.62	0.82
RT_Making	-0.54***	0.15	-3.64			
Resp.Length	-0.21***	0.05	-4.47	-0.22***	0.05	-4.10
Ort.Similarity	-8.28***	1.74	-4.78	-5.98***	1.79	-3.34
Sem.Similarity.p3				0.11***	0.03	4.35
Ort.Similarity.p7				11.45***	1.81	6.33

[1]p-value significance codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1

Table 4. Random Effects (GLMM).

	Model1 (RTV)		Model2 (RTV + CTV)		
	Var	StDev	Var	StDev	Corr.
Target (Intercept)	0.48	0.69	1.05	1.02	
Familar – Unknown:Known			2.47	1.57	-1.00
Familiar – Unknown:Familiar			23.00	4.80	-1.00
Subject(Intercept)	4.97	2.23	3.67	1.92	

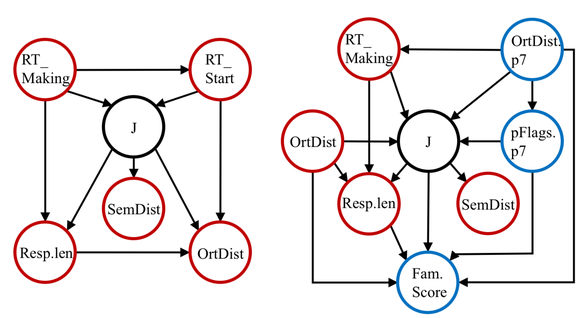
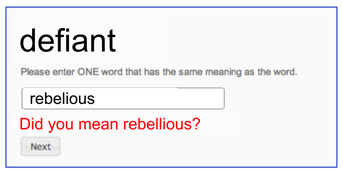


Figure 4. Suggested interaction structure of RTV and R+CTV models from the Hill-climbing algorithm
(Off-task label: black node (J), RTVs: red nodes, CTVs: blue nodes).

Figure 2. An example of the meaning-generation task with incorrectly spelled response (left).All responses are recorded and provided to human judges to label off-task responses within the student (middle) and across students in a similar time frame (right).



j1	j2	Student	Seq #	Trial	Targ	Response
*		S1	25	1	nostalgia	ocean
*		S1	39	1	turmoil	maybe
*	*	S1	47	1	assimilate	One
*	*	S1	51	1	caustic	Two

j1	j2	Student	Seq #	Trial	Targ	Response
		S14	11	1	reticent	receive
		S7	12	1	perturbed	clean
*	*	S26	13	1	tenable	rain
*	*	S22	14	1	vie	Rain

CONTRIBUTION

❑ METHODS FOR EXTRACTING FEATURES

- Methods for **extracting meaningful information** from log data.
- **RTV + CTV** with mixed effect model:
  - **CTVs** improve the model performance.
  - **CTVs** can **substitute traditional off-task predictive features**, such as # of error messages and response time.

❑ IDENTIFYING OFF-TASK STATUS AT THE ITEM LEVEL

- Letting the learning system know **when to intervene** with students to retain engagement.
- Manage student **engagement** systematically.

❑ MORE ACCURATE PREDICTION ON THE STUDENT'S VOCABULARY KNOWLEDGE

- Distinguishing between accidentally erroneous responses and intentionally missed responses.

FUTURE WORK

❑ DEVELOP AN ADAPTIVE VOCABULARY LEARNING SYSTEM

- Adaptive system that can **minimize the off-task** behaviors during the learning task.
- Find **desirable difficulty** level for each student.
- **Identify behavioral log features** related with **perceived difficulty**.

❑ LABELING FROM NON-EXPERTS

- Fragmentary job **for anonymous workers**.
- Require more careful design instructions.

References

[1] R. S. Baker, A. T. Corbett, and K. R. Koedinger. Detecting student misuse of intelligent tutoring systems. In *Intelligent tutoring systems*, pages 531–540. Springer, 2004.

[2] Frishkoff, G., Collins–Thompson, K., Nam, S.J., Hodges, L., & Crossley, S. (in press). Dynamic Support of Contextual Vocabulary Acquisition for Reading (DSCoVAR): An intelligent tutoring system for contextual word learning. In S.A. Crossley & D.S. McNamara (Eds.), *Handbook on Educational Technologies for Literacy,* Taylor and Francis.

[3] Collins-Thompson, K., Frishkoff, G., & Crossley, S.A. (2012). Definition Response Scoring with Probabilistic Ordinal Regression. *Proceedings of the 20th International Conference on Computers in Education (ICCE 2012*).

[4] Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, 15(1), 71–91.

[5] S. Nam, K. Collins-Thompson, and G. Frishkoff. Modeling real-time performance on a meaning-generation task. In *Annual Meeting of the American Educational Research Association*. AERA, 2016.

[6] S. Nam, K. Collins-Thompson, G. Frishkoff, and L. Hodges. Measuring real-time student engagement in contextual word learning. In *The 22nd Annual Meeting of the Society for the Scientific Study of Reading*. SSSR, 2015. https://goo.gl/CvTL1K.

Acknowledgments

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, Grant R305A140647 (co-PI Collins-Thompson & Frishkoff). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We thank Falk Laboratory School for their participation. We also thank Russell Brand, Marika Hein, Junaid Amin, Austin Marcus and Goda Tarcijonas for assistance with the tutor and stimulus development.