Off-Task Behavior in Narrative-Centered Learning Environments

Jonathan P. ROWE[a, 1], Scott W. MCQUIGGAN[b], Jennifer L. ROBISON[a], and James C. LESTER[a]
[a] *Department of Computer Science, North Carolina State University, Raleigh NC 27695*
[b] *Education Practice, SAS Institute, Inc., Cary NC 27513*

Abstract. Recent years have seen increasing interest in narrative-centered learning environments. However, the same qualities that make them engaging can also introduce seductive details that invite off-task behavior. This paper examines off-task behavior in the CRYSTAL ISLAND narrative-centered learning environment. Results from an empirical study examining the relationships between student test performance, individual differences, and off-task behavior are presented. The study found negative correlations between off-task behavior and test performance, as well as significant gender effects on the total amount of off-task behavior. Initial conclusions from a path analysis conducted on students' action sequences are also presented.

Keywords. Game-based learning environments, Narrative-centered learning environments, Off-task behavior

1. Introduction

Narrative-centered learning environments (NLEs) have become the subject of increasing attention in the Artificial Intelligence in Education community [1, 2, 3, 4]. By contextualizing learning within narrative settings, NLEs tap into students' innate facilities for crafting and understanding stories [5, 6], and they take advantage of narratives' motivating features such as compelling plots, engaging characters, and fantastical settings [7]. NLEs offer significant potential for encouraging active participation in learning, higher-level thinking, and forming connections between narrative and pedagogical content. By embedding technologies from intelligent tutoring systems, embodied conversational agents, and serious games into narrative-centered virtual environments, NLEs offer the promise of adaptive, situated learning experiences that are highly interactive and engaging for students. NLEs are currently under investigation in a range of domains, including language learning [3], anti-bullying education [2], health intervention education [1], and science learning [4].

While narrative-centered learning environments offer several attractive qualities, these same characteristics can invite students to participate in off-task behaviors. Educators have long striven to reduce off-task behavior in classrooms and intelligent tutoring systems [8, 9, 10, 11, 12]. In these settings, student disengagement has

[1] Corresponding Author: Jonathan P. Rowe, Department of Computer Science, North Carolina State University, Raleigh, NC 27695; E-mail: jprowe@ncsu.edu

commonly taken the form of off-topic conversation, "gaming the system" behaviors, participation in non-relevant side activities, or general inactivity. Unfortunately, NLEs often provide vast interactive environments, realistic physics, and engaging characters that introduce additional *seductive details* to learning experiences. Seductive details have the potential to distract, disrupt, or divert students' attention from pedagogical objectives and to reduce students' time-on-task [13]. In NLEs, off-task behavior and on-task behavior can appear similar to each other—to the casual observer the student is simply playing the game—thereby presenting an additional challenge to instructors. While the motivational and pedagogical benefits of NLEs are compelling, it is important for researchers and educators to develop a clear understanding of off-task behavior so that they can address any adverse effects on learning.

In this paper, we examine students' off-task behavior in the CRYSTAL ISLAND narrative-centered learning environment. We outline the relationship between others' work on off-task behavior in ITSs and off-task behavior in CRYSTAL ISLAND. We then present results from an empirical study examining individual differences in off-task behavior, and their relationships to students' test scores. We also summarize initial conclusions from a path analysis conducted on students' action sequences for completing CRYSTAL ISLAND's objectives. We conclude with a discussion of implications for the future development of NLEs and directions for subsequent work.

2. Off-task Behavior in Learning Environments

Off-task behavior is a symptom of disengagement from a learning experience. High levels of engagement are important for maintaining student motivation and maximizing students' time-on-task. A number of automated approaches for diagnosing student engagement in ITSs have been proposed by AI researchers, including hidden Markov models [14], item response theory-based approaches [15], and other machine-learning-based techniques [9]. Constructs such as motivation [7, 16, 17] and presence [4, 18] have also been investigated in related work. However, unlike many other learning technologies in which interacting in the "normal, expected" manner suggests that the student is on-task, off-task behavior in NLEs can also be manifested as "normal, expected" interactions within the software. Off-task behavior in NLEs consists of the student disengaging from pedagogically relevant behavior, i.e., students can attentively interact with the NLE but actually be focused on activities with little intrinsic educational value. For these reasons, we cannot solely rely upon students' appearance of being engaged in the software as a reliable indicator of being off-task: we must also attend to their behavior in the virtual environment.

One of the most frequently studied types of in-software off-task behavior by the AI in Education community is "gaming the system" [8, 9, 10, 12, 19]. Gaming behavior has been shown to have a particularly negative impact on learning, and to correlate with students' lack of self-drive, frustration, and dislike of subject matter [19]. While gaming is an important type of off-task behavior in both traditional ITSs and NLEs, we focus on a narrower definition of off-task behavior for the purposes of this work.

Our work on off-task behavior is situated in CRYSTAL ISLAND, a narrative-centered learning environment built on Valve Software's Source™ engine, the 3D game platform for Half-Life 2. CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island. Students play the role of the protagonist, Alyx, who is attempting to discover the identity and source of an unidentified infectious disease plaguing a newly established research station. Typical game play involves navigating the island, manipulating objects, taking notes, viewing posters, operating lab equipment, and talking with non-player characters to gather clues about the disease's source. CRYSTAL ISLAND's setting includes a beach area with docks, a large outdoor field laboratory, underground caves, and a research camp. To progress through the mystery, students must explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses.

The notion of off-task behavior investigated here targets the frequency and duration of students' attendance to purely narrative features rather than pedagogically useful aspects of CRYSTAL ISLAND. In the seminal work on seductive details in expository texts, the phenomenon is defined as "highly interesting and entertaining information that is only tangentially related to the topic but is irrelevant to the author's intended theme" [13]. Narrative-centered learning environments, and other game-based learning tools, frequently introduce elements that are extraneous to core subject matter in the form of interactive gameplay elements, locations, or narrative features for purposes of engagement. Elements such as these can be viewed as a form of seductive details. Building on this notion of seductive details as they occur in narrative-centered learning, we define any student behavior that involves locations or objects not necessary for solving CRYSTAL ISLAND'S science mystery as off-task. In doing so, we adopt a relatively conservative definition of off-task behavior: only behaviors that are clearly unrelated to the narrative and curriculum are denoted as off-task. This definition represents just a first step toward automatically classifying off-task behavior in NLEs. In the following sections, we explore how this notion of off-task behavior correlates with student test scores and related variables in an empirical study of CRYSTAL ISLAND.

3. Method

3.1. Participants

There were 88 female and 91 male participants varying in age and race. Approximately 2% of the participants were American Indian or Alaska Native, 5% were Asian, 29% were Black or African American, 58% were Caucasian, 6% were Hispanic or Latino, and 6% were of other races. Participants were all eighth-grade students ranging in age from 12 to 15 (M = 13.27, SD = 0.51). The students had recently completed the microbiology curriculum mandated by the North Carolina state standard course of study before receiving the instruments, tests, and interventions of this experiment.

3.2. Materials and Apparatus

The pre-experiment paper-and-pencil materials for each participant were completed one week prior to intervention. These materials consisted of a researcher generated CRYSTAL ISLAND curriculum test, demographic survey, Achievement Goals Questionnaire [20], Self-Efficacy for Self-Regulated Learning scale (SESRL) [21], Science Self-Efficacy scale, modified from [22], and Immersion Tendencies Questionnaire [23]. The CRYSTAL ISLAND curriculum test consists of 23 questions created by an interdisciplinary team of researchers and was approved for language and content by the students' eighth-grade science teachers.

Post-experiment materials were completed immediately following intervention. These materials consisted of the same CRYSTAL ISLAND curriculum test, Achievement Goals Questionnaire [20], Science Self-Efficacy scale, an interest scale [24], and the Presence Questionnaire [23]. The interest scale was adapted from those used by Schraw to capture differences across groups and to examine within-subject relationships with learning outcomes [19]. Participants' presence experience was captured with the Presence Questionnaire (PQ) developed and validated by Witmer and Singer [21].

4. Design and Procedure

4.1. Design

The experiment randomly assigned the entire eighth grade population of a middle school in Raleigh, North Carolina to four groups: holdout, CRYSTAL ISLAND narrative condition, CRYSTAL ISLAND minimal-narrative condition, or PowerPoint condition. Participants in the holdout condition did not receive an intervention and served as the control group for this experiment and planned longitudinal studies. In the remaining three conditions, students were exposed to the CRYSTAL ISLAND microbiology curriculum delivered in one of three formats. Students in the PowerPoint condition were presented with a narrated slide show of the curriculum content. They did not play a version of CRYSTAL ISLAND at the time of the study, and their data is not considered in this analysis. The CRYSTAL ISLAND narrative condition supplemented the curriculum with the full CRYSTAL ISLAND narrative, including a poisoning scenario, character back-stories, and rich character personalities. The CRYSTAL ISLAND minimal-narrative condition supplemented microbiology content with the minimal story required to support the curriculum.

4.2. Participant Procedure

Participants entered the experiment room having completed the pre-test and instrumentation one week prior to the intervention. Participants were first instructed to review CRYSTAL ISLAND instruction materials. These materials consisted of the CRYSTAL ISLAND back-story and task description, a character handout, a map of the island, and a control sheet. Participants were then further directed on the use of the controls via a presentation explaining each control in detail.

Participants in the three intervention conditions (narrative, minimal-narrative, and PowerPoint) were given 50 minutes to work. Solving the mystery in the narrative and minimal-narrative conditions consisted of completing a number of goals including learning about pathogens, viruses, bacteria, fungi, and parasites, compiling the symptoms of the researchers who had fallen ill, recording features of hypothesized diseases causing the CRYSTAL ISLAND illness, testing a variety of possible sources, and reporting the solution (cause and source) to the camp nurse to design a treatment plan.

Immediately after solving the science mystery, or 50 minutes of interaction, participants completed the post-experiment questionnaires. First to be completed was the CRYSTAL ISLAND curriculum test, followed by the remaining post-experiment questionnaires described above. Completion of post-experiment materials took no

Table 1. CRYSTAL ISLAND location frequencies correlated with post-test scores.
○ – Denotes a location included in the off-task behavior definition.

Location	Pre-test score correlation			Post-test score correlation		
(mean freq.)	*r*(92)=	*p*		*r*(92)=	*p*	
○ **Waterfall,** *M*=1.1, *SD*=1.1	-0.3626	<.0001	*	-0.4341	<.0001	*
Beach, *M*=6.8, *SD*=4.2	-0.3469	0.001	*	-0.2600	0.014	*
○ **Cave,** *M*=1.4, *SD*=1.3	-0.3400	0.001	*	-0.1833	0.084	
Path, *M*=7.9, *SD*=3.6	-0.2783	0.008	*	-0.1381	0.194	
Hilltop, *M*=2.7, *SD*=1.8	-0.1854	0.080		-0.1153	0.279	
○ **Restrooms,** *M*=0.7, *SD*=0.6	-0.1764	0.096		-0.2037	0.054	
Docks, *M*=1.7, *SD*=2.0	-0.1629	0.125		-6.1700	0.108	
Women's Quarters, *M*=2.1, *SD*=1.3	0.0615	0.564		0.1536	0.140	
Bryce's House, *M*=1.6, *SD*=1.2	0.0642	0.547		0.1682	0.113	
Laboratory, *M*=2.9, *SD*=2.5	0.1055	0.322		0.2030	0.050	*
Dining Hall, *M*-2.9, *SD*=2.2	0.1422	0.181		0.2473	0.019	*
○ **Water**, *M*=2.4, *SD*=2.3	0.1734	0.102		-0.1973	0.062	
Infirmary, *M*=3.4, *SD*=1.9	0.2480	0.018	*	0.4638	<.0001	*
Men's Quarters, *M*=3.9, *SD*=1.9	0.2757	0.009	*	0.4540	<.0001	*

longer than 35 minutes for participants. In total, sessions lasted 90 minutes.

5. Results

Significant learning gains were found in both conditions (a discussion of these results may be found in [4]), indicating that students did learn by playing CRYSTAL ISLAND. During these interactions, students engaged in off-task behavior on average 14.98% of the time (SD = 8.9%). The proportion of off-task behavior negatively correlated with both pre-test score, $r(92) = -0.3455$, $p = 0.0007$, and post-test score, $r(92) = -0.3740$, $p = 0.0002$. The correlation with learning gains as measured by the difference between pre and post test scores was not statistically significant, $r(92) = -0.01$, p=0.92. Investigating the locations identified in our conservative definition of off-task behavior, we find interesting relationships with test performance. Table 1 correlates test scores with frequencies of visits to CRYSTAL ISLAND locations. Interestingly, all frequencies of off-task location visits have, at least, weakly significant ($p < .1$) correlations with post-test scores. The remaining locations are pertinent to narrative or learning goals in CRYSTAL ISLAND. Experimental condition (narrative vs. minimal-narrative) had no effect on the frequency of student off-task behaviors, $t(92) = 0.75$, $p = 0.46$, with students in the narrative condition ($M = 100.4$, $SD = 8.65$) engaging in approximately nine more off-task behaviors than students in the minimal-narrative condition ($M = 91.4$, $SD = 8.29$) on average.

Overall, male students engaged in significantly more off-task behaviors (nearly twice as much) $t(92) = 5.37$, $p < 0.0001$. Male students had a mean frequency of off-task behaviors of 120.5 (*SD* = 6.9) while females had a mean frequency of 63.7 off-task behaviors (*SD* = 7.9) while solving the CRYSTAL ISLAND mystery. No significant relationship was found between off-task behavior and self-efficacy, goal orientation, interest, or presence. Results indicate that off-task behavior had no effect on how student's enjoyed the learning experience, their confidence in their ability to solve microbiology problems, nor their desire to play CRYSTAL ISLAND again.

6. Discussion

The results illustrate that off-task behavior has a conclusive, negative relationship with pre and post test performance, and that students' gender bears on their off-task behaviors. Building upon these results, further work will be necessary to establish what additional student characteristics correlate with off-task behavior in NLEs, and what concrete design implications should be drawn as a consequence. For example, the authors' informal observations suggest that some students may unknowingly spend time in off-task locations searching for useful information that is not there, whereas others may simply be off-task. Further, some NLE locations may be engaging but elicit off-task behaviors. For example, students' interaction with CRYSTAL ISLAND's `Beach` location illustrates such a situation; it was found to have a negative correlation with students' test scores, and it contains only peripheral information relevant to the story of CRYSTAL ISLAND—no curricular goals can be achieved there. It also seems feasible that purely narrative elements may introduce tradeoffs between engagement and learning for different students, and students' individual differences may impact the nature of these tradeoffs.

It is notable that the post test results and overall proportion of time spent off-task resemble other results collected in a substantially different learning environment (a cognitive tutor) using substantially different definitions for off-task behaviors [8]. The lack of relationship between performance orientation and off-task behavior resemble the cognitive tutor results, as well [19]. Because the proportion of off-task behavior recorded was relatively low, it seems promising that further research may determine whether some off-task behavior can be deemed desirable as a short, engaging break from learning activities in NLEs.

An important next step in the research agenda is designing intervention strategies to scaffold off-task students. Using analysis techniques common among web analytics, we have conducted a path analysis (Figure 1) of CRYSTAL ISLAND using SAS® Enterprise Miner™. The path analysis includes a frequency analysis, calculating the frequency of visited locations as well as the frequency of paths taken in the CRYSTAL ISLAND environment. Using the link graph, we can identify candidate locations for pedagogical intervention in hopes of managing off-task behavior and encouraging routes that lead to learning objectives. For example, the majority of CRYSTAL ISLAND's curriculum is encountered through problem-solving activities in the camp area (dining hall, infirmary, etc.). Providing a pedagogical agent near the `Path` location may be a useful tactic for encouraging students to remain on-task. Further, link graphs can encode useful information to be leveraged by pedagogical agents and companion agents. A link graph can be used to determine what paths a student has taken, as well as paths taken by successful students for the same activity, and then combine this information to infer what types of context appropriate hints might be most effective.

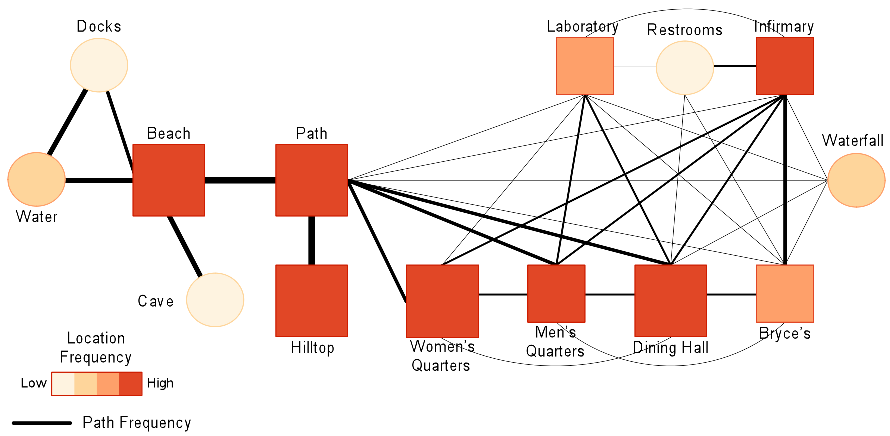

Figure 1. Path analysis link graph from the narrative experimental condition.

7. Conclusions

This work represents an initial step toward a more comprehensive understanding of off-task behavior in narrative-centered learning environments. As evidenced by a sample of empirical results, off-task behavior in narrative-centered learning environments exhibit both notable similarities and notable differences with respect to traditional intelligent tutoring systems. As narrative-centered learning environments and other game-based learning environments are more widely adopted, it will become increasingly important to develop a methodology for understanding how students play them, and what role seductive gameplay features, targeted at increasing engagement, should have in these learning environments. Initial results indicate that unnecessary narrative details may be particularly seductive for particular types of students, such as lower-performing students and male students.

Moving forward, these results suggest the need to develop more sophisticated notions of off-task behavior. Using them as a springboard, we can then induce models for automatically detecting harmful in-game behaviors, and engineer adaptive scaffolding techniques for re-directing students toward more pedagogically beneficial activities.

Acknowledgements

The authors wish to thank members of the IntelliMedia lab for their assistance, Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the Source™ engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, IIS-0757535, DRL-0822200 and IIS-0812291. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Marsella, W. L. Johnson and C. LaBore, Interactive pedagogical drama for health interventions, in *Proceedings of the 11th International Conference on Artificial Intelligence in Education* (2003), 341-348.

[2] R. Aylett, S. Louchart, J. Dias, A. Paiva and M. Vala, FearNot! an experiment in emergent narrative, in *Proceedings of the 5th International Conference on Intelligent Virtual Agents* (2005), 305-316.

[3] W. L. Johnson, Serious use of a serious game for language learning, in *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (2007), 67-74.

[4] S. McQuiggan, J. Rowe, S. Lee and J. Lester, Story-based learning: the impact of narrative on learning experiences and outcomes, in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (2008), 530-539.

[5] C. Wells, *The meaning makers: children learning language and using language to learn*, Heinemann, Portsmouth, NH, 1986.

[6] J. Bruner, The narrative construction of reality, *Critical Inquiry* **18** (1991) 1-21.

[7] T. Malone and M. Lepper, Making learning fun: a taxonomy of intrinsic motivations for learning, in R. Snow and M. Farr (Eds.) *Aptitude, learning, and instruction: III. Conative and affective process analyses,* (pp. 223-253), Erlbaum, Hillsdale, NJ, 1987.

[8] R., Baker, A. Corbett, K. Koedinger and A. Wagner, Off-task behavior in the cognitive tutor classroom: when students "game the system", in *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2004), 383-390.

[9] J. Walonoski and N. Heffernan, Detection and analysis of off-task gaming behavior in intelligent tutoring systems, in *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (2006), 382-391.

[10] I. Arroyo, K. Ferguson, J. Johns, T. Dragon, H. Mehranian, D. Fisher, A. Barto, S. Mahadevan and B. Woolf, Repairing disengagement with non-invasive interventions, in *Proceedings of the 13th International Conference of Artificial Intelligence in Education* (2007), 195-202.

[11] R. Baker, Modeling and understanding students' off-task behavior in intelligent tutoring systems, in *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2007), 1059-1068.

[12] M. Rodrigo, R. Baker, M. Lagud, S. Lim, A. Macapanpan, S. Pascua, J. Santillano, L. Sevilla, J. Sugay, S. Tep and N. Viehland, Affect and usage choices in simulation problem solving environments, in *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (2007), 145-152.

[13] S. Harp and R. Mayer, How seductive details do their damage: a theory of cognitive interest in science learning, *Journal of Educational Psychology* **90**(3) (1998), 414-434.

[14] C. Beal, S. Mitra and P. Cohen, Modeling learning patterns of students with a tutoring system using hidden markov models, in *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (2007), 238-245.

[15] J. Beck, Engagement tracing: using response times to model student disengagement, in *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (2005), 88-95.

[16] A. Vicente and H. Pain, Informing the detection of the student's motivational state: an empirical study, in *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* (2002), 933-943.

[17] J. Johns and B. Woolf, A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the 21st National Conference on Artificial Intelligence* (2006).

[18] S. McQuiggan, J. Rowe and J. Lester, The effects of empathetic virtual characters on presence in narrative-centered learning environments, in *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2008), 1511-1520.

[19] R. Baker, J. Walonoski, N. Heffernan, I. Roll, A. Corbett and K. Koedinger, Why students engage in "gaming the system" behavior in interactive learning environments, *Journal of Interactive Learning Research* **19**(2) (2008), 185-224.

[20] A. Elliot and H. McGregor, A 2 x 2 achievement goal framework, *Journal of Personality and Social Psychology* **80**(3) (2001), 501-519.

[21] A. Bandura, Guide for constructing self-efficacy scales, in F. Pajares and T. Urdan (Eds.) *Self-Efficacy Beliefs of Adolescents,* 307-337, Information Age Publishing, Greenwich, Connecticut, 2006.

[22] J. L. Nietfeld, L. Cao and J. W. Osborne, The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy, *Metacognition and Learning* **1**(2) (2006), 159-179.

[23] B. Witmer and M. Singer, Measuring presence in virtual environments: a presence questionnaire, *Presence: Teleoperators and Virtual Environments* **7**(3) (1998), 225-240.

[24] G. Schraw, Situational interest in literary text, *Contemporary Educational Psychology* **22** (1997), 436–456.