Investigation of the Influence of Hint Type on Problem Solving Behavior in a Logic Proof Tutor

Christa Cody(✉), Behrooz Mostafavi, and Tiffany Barnes

North Carolina State University, Raleigh, USA
{cncody,bzmostaf,tmbarnes}@ncsu.edu

Abstract. Within intelligent tutoring systems, hint policies are needed to determine when and how to give hints and what type of hint is most beneficial. In this study, we focus on discovering whether certain hint types influence problem solving behavior. We investigate the influence of two hint types (next-step hints and more abstract high-level hints) on students' behavior in a college-level logic proof tutor, Deep Thought. The results suggest that hint types can affect student behavior, including hint usage, rule applications, and time in-tutor.

Keywords: Tutoring system · Hint type · Data-driven

1 Introduction

Intelligent tutoring systems (ITS) provide adaptive instruction to students and have significant effects on learning [1]. Data-driven methods, where actions within the tutor are based on analyzing historical data, have been used to great effect to individualize computer-aided instruction [2–4]. Within an ITS, hint policies are needed to determine when and how to give hints and what type of hint is most beneficial [5]. The most minimal hint type is error-specific feedback, which provides a hint regarding an error the student has made [5]. However, research has suggested that goal-directed feedback leads to better performance by providing students with direction in solving the problem [6]. A common approach is to provide a sequence of hints beginning with a more general hint then transition to more specific and directive hints (e.g. Point, Teach, and Bottom-out [7]). However, students may benefit from receiving a certain hint type over another, so providing hints in a strict sequence might limit the effectiveness of receiving hints. Some research has created individualized hint policies based on student behavior and ability, allowing the student to receive the appropriate hint without stepping through strict levels of hints [8,9]. Additionally, research has shown that high ability learners benefit from lower amounts or more abstract guidance while low ability learners benefit from higher amounts or more concrete guidance [10,11]. The goal of this work is to determine if hint type influences student's performance and behavior.

C. Penstein Rosé et al. (Eds.): AIED 2018, LNAI 10948, pp. 58–62, 2018.
https://doi.org/10.1007/978-3-319-93846-2_11

2 Method

We tested our two hint types in a propositional logic proof tutor, Deep Thought [2,12], as assigned homework in a two sections of a college-level discrete mathematics course taught by the same instructor in Fall 2016. The tutor presents worked examples and problems consisting of logical premises and a conclusion to be derived using logical axioms, and is divided into 6 levels. At the end of each level, the student is presented with a final problem with no hints, which serves as a level posttest. Hint type was randomly assigned for Next Step Hints (NSH, n = 48) or High Level Hints (HLH, n = 47) before tutor use. Students who collaborated or dropped out of the tutor were removed (NSH n = 12, HLH n = 11), because their data would skew time and step features. Therefore, the following analyses were based on data from 36 NSH and 36 HLH students. To analyze the specific influences hint type has on student problem solving, each condition provided one of two hint types: **Next-Step Hints (NSH)** or **High-Level Hints (HLH)**. Next-step hints suggest the next step of a logic proof that can be immediately implemented in the student's current proof – providing more explicit instruction. Whereas, high-level hints represent hints that can be 2 or 3 possible rule applications ahead of the current proof state [13]. High-level hints are aimed at helping the student develop a strategy. There is a significant problem regarding help avoidance where students may avoid asking for hints even when they might need them [14], so we implemented proactive hints by providing a hint when the student is taking longer than the median time to correctly complete a single step.

2.1 Description of Metrics

To examine the effects of hint types, we focused on behavior- and performance-related features. In Level 1, all students receive the same set of problems and are not given any worked examples. Therefore, the student's performance on Level 1 is used as a **pretest** to measure their incoming knowledge. The average performance of Levels 3–6 is used as a **posttest** to measure their post-training performance. Additionally, all students were evaluated using two proof problem questions as part of a mid-term examination. **Total Correct/Incorrect Steps** are the number of correct/incorrect rule applications over the course of the whole tutor. **Mean/Median Correct Step Time** and **Mean/Median Incorrect Step Time** represent the mean/median time between steps (correct and incorrect, respectively) for each student. **Total Time to Complete** is the total time spent solving a problem in minutes. **Final Solution Length** is the number of derivations made to reach the conclusion in the final solution. We also analyzed the number of hints requested, proactive hints, and total hints received.

The data is positively skewed so we transformed it using a log transform then applied ANOVA to find significant differences between the groups. We examined the metrics for the end of level problems (containing no hints) to discover differences in problem solving that may be affected by the hint type.

3 Results and Discussion

First, we looked at the performance metrics, which did not show any significant differences between the groups. Next, we analyzed the differences between students behavior using the end of level problems, which allows a view of how the hints may be shaping the behavior of a student's problem solving. For the sake of space, we selected representative problems and only report behaviors that are indicative of differences between the two groups. In Table 1, the HLH group took significantly less time solving the problem and with significantly less incorrect steps. However, the HLH group spent approximately the same amount of time ($p = 0.641$) on Level 1, which means that they spent longer during the initial few problems before speeding up in the end of level problem.

Table 1. Stats and significant ANOVA results for the end of level problem for Level 1

End of level problem - Level 1					
Metric	NSH (n = 36)		HLH (n = 36)		
	Mean(SD)	Median	Mean(SD)	Median	p-value
Total incorrect steps	5(5)	1	2(2)	0	0.296
Total correct steps	6(5)	4	5(1)	4	0.200
Time to complete (min)	7.3(12)	2.3	3.2(4.2)	1.6	.0365*
Median correct step time (s)	63(122)	26	34(39)	23	0.067

In Table 2, the NSH group has significantly more correct steps ($p = 0.007$) and marginally more incorrect steps ($p = 0.062$). Due to the nature of the tutor and solving proofs, students may make as many derivations as they want; however, having higher correct and incorrect steps is indicative of gaming behavior. The NSH group also spent significantly longer on this problem ($p = 0.002$).

Table 2. Stats and significant ANOVA results for the end of level problem for Level 3

End of level problem - Level 3					
Metric	NSH (n = 36)		HLH (n = 36)		
	Mean(SD)	Median	Mean(SD)	Median	p-value
Total incorrect steps	16(25)	6	5(7)	2	0.062
Total correct steps	13(9)	8	9(3)	7	0.007*
Time to complete (min)	26.2(69)	4	4.5(4)	3	0.002*
Mean incorrect step time (s)	80(101)	45	48(29)	37	0.089
Mean correct step time (s)	72(115)	29	33(33)	23	0.015*

In Table 3, the NSH had significantly more incorrect steps ($p = 0.042$), took longer to complete the problem ($p = 0.021$), took longer to correctly apply rules (p = 0.042), and had a longer final solution (p = 0.058). These metrics indicate that the NSH group has less knowledge or strategy to solve the problem as compared to the HLH group.

Table 3. Stats and significant ANOVA results for the end of level problem for Level 5

End of level problem - Level 5					
Metric	NSH (n = 36)		HLH (n = 36)		
	Mean(SD)	Median	Mean(SD)	Median	p-value
Total incorrect steps	24.05(44)	8.5	11.2(14)	5	0.042*
Total correct steps	20.15(18)	13.5	16.15(12)	12.5	0.687
Time to complete (min)	36.1(60)	11.5	13.2(22.8)	6.7	0.021*
Mean incorrect step time (s)	466(2060)	54.4	61(64)	43.6	0.066
Mean correct step time (s)	124(262)	33.7	37(24)	28.6	0.042*
Final solution length	10.5(3)	10	9.5(2)	9	0.058*

During the course of the tutor, there is a widening gap between NSH and HLH group with respect to the time it takes to complete the problem, the amount of incorrect steps made, and final solution length. One possibility for this difference in behavior is HLH are promoting strategy-oriented behavior, which, if so, we would expect the students to make fewer mistakes and less steps because they would be planning out a few steps ahead thus less likely to go down an incorrect path. Another possibility is the NSH students could have been using the hints to directly discover what to do next, which could have led to gaming behavior [15].

4 Conclusion and Future Work

In this study, we compared two test conditions of Deep Thought, HLH and NSH hint groups, to further the understanding of the impact of hint type on the behavior of students during problem solving. We found that students receiving HLH seem to produce more positive in-tutor behavior resulting in less incorrect steps and less time spent during end of level problems, while students in the NSH group seemed to have less desirable in-tutor behavior, including larger amounts of incorrect steps. Even though there were no significant differences in the performance metrics, longer time spent solving a problem and learned strategies are a large concern due to possible transfer of behavior outside of the tutor. Large amounts of incorrect steps, especially occurring closer together, are indicative of gaming the tutor to finish a proof.

To further explore the influences of hint type, we will be conducting a more in-depth study of the behavior surrounding hints, including how long it took the student to derive a hint, and what type of behavior indicates when each type of hint is beneficial. This can be further analyzed by looking into how and which hints are followed. A more formal posttest will be used in future evaluations to provide more insight to the students' learning over the course of the tutor.

References

1. Murray, T.: Authoring intelligent tutoring systems: an analysis of the state of the art. Int. J. Artif. Intell. Educ. **10**, 98–129 (1999)
2. Mostafavi, B., Barnes, T.: Evolution of an intelligent deductive logic tutor using data-driven elements. Int. J. Artif. Intell. Educ. **27**, 1–32 (2016)
3. Stamper, J., Eagle, M., Barnes, T., Croy, M.: Experimental evaluation of automatic hint generation for a logic tutor. Int. J. Artif. Intell. Educ. **22**(1–2), 3–17 (2013)
4. Fossati, D., Di Eugenio, B., Ohlsson, S., Brown, C., Chen, L.: Data driven automatic feedback generation in the ilist intelligent tutoring system. Technol. Instr. Cogn. Learn. **10**(1), 5–26 (2015)
5. Vanlehn, K.: The behavior of tutoring systems. Int. J. Artif. Intell. Educ. **16**(3), 227–265 (2006)
6. McKendree, J.: Effective feedback content for tutoring complex skills. Hum.-Comput. Interact. **5**(4), 381–413 (1990)
7. Hume, G., Michael, J., Rovick, A., Evens, M.: Hinting as a tactic in one-on-one tutoring. J. Learn. Sci. **5**(1), 23–47 (1996)
8. Stern, M., Beck, J., Woolf, B.P.: Adaptation of problem presentation and feedback in an intelligent mathematics tutor. In: Frasson, C., Gauthier, G., Lesgold, A. (eds.) ITS 1996. LNCS, vol. 1086, pp. 605–613. Springer, Heidelberg (1996). https://doi.org/10.1007/3-540-61327-7_160
9. Anohina, A.: Advances in intelligent tutoring systems: problem-solving modes and model of hints. Int. J. Comput. Commun. Control **2**(1), 48–55 (2007)
10. Arroyo, I., Beck, J.E., Woolf, B.P., Beal, C.R., Schultz, K.: Macroadapting animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In: Gauthier, G., Frasson, C., VanLehn, K. (eds.) ITS 2000. LNCS, vol. 1839, pp. 574–583. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45108-0_61
11. Luckin, R., Du Boulay, B., et al.: Ecolab: the development and evaluation of a Vygotskian design framework. Int. J. Artif. Intell. Educ. **10**(2), 198–220 (1999)
12. Mostafavi, B., Eagle, M., Barnes, T.: Towards data-driven mastery learning. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp. 270–274. ACM (2015)
13. Eagle, M., Johnson, M., Barnes, T.: Interaction networks: generating high level hints based on network community clustering. International Educational Data Mining Society (2012)
14. Aleven, V., Koedinger, K.R.: Limitations of student control: do students know when they need help? In: Gauthier, G., Frasson, C., VanLehn, K. (eds.) ITS 2000. LNCS, vol. 1839, pp. 292–303. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45108-0_33
15. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting student misuse of intelligent tutoring systems. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 531–540. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30139-4_50