HELP-SEEKING IN AN ONLINE MATHS ENVIRONMENT: A SEQUENCE ANALYSIS OF LOG FILES

Christian Bokhove

University of Southampton; C.Bokhove@soton.ac.uk

In online mathematics environments feedback is often provided to help students make progress. However, the extent to which students make use of such feedback, so-called 'help-seeking', depends on numerous instructional variables, including the design of the online platform and individual student characteristics. Furthermore, student behaviour in such platforms are not independent events: the order in which tasks are completed matters, and we therefore have to study sequences of such events. This study used student data from UK students in grades 3 to 5 with at least 100 lesson records in the academic year '18-'19 (N=1,799), totalling 1,048,575 records between December 2010 and January 2019, from an online mathematics platform. Sequence analysis was applied to the data to uncover patterns of help-seeking. The results showed that help, task difficulty and precision interact, demonstrating the usefulness of learning analytical approaches like sequence analysis.

Keywords: help-seeking, online mathematics platform, sequence analysis, feedback

BACKGROUND

The recent global pandemic has shown how in a relatively short time, online provision has become essential in the learning experiences of school pupils. This, however, is not without its challenges. Utilising computers to aid tutoring is a relatively recent development in education and coincides with the development of computers. Whereas human tutoring has been used in schools for 2,500 years—or for as long as schools have existed—computer tutoring is largely a product of the past half century (Kulik & Fletcher, 2016). In this period numerous reviews have appeared that tried to summarise the efficacy of such technological developments. Already, in the 70s and 80s computer tutoring systems appeared that tried to combine artificial intelligence concepts and cognitive theory. Such systems guided learners through each step of a problem solution by giving hints and feedback. The first-generation computer tutors are now often referred to as CAI tutors (computer-assisted instruction), the more modern variations are usually called Intelligent Tutoring Systems, or ITSs (VanLehn, 2011), which distinguish themselves from CAI tutors because they can, in addition to providing feedback on student answers, also provide feedback on the thinking that goes into individual answers. ITSs in that sense are systems that present feedback and/or hints (but they in essence also are feedback) to learners as to assist the learning process. For Mousavinasab et al. (2018) this also includes adaptive guidance and instruction, which I take to be guidance that can vary depending on the learner's context. As reported by Du Boulay (2016), there have been a number of recent positive reviews in support of the effectiveness of ITSs (Kulik & Fletcher, 2016; Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011). Thus, it is well known that well-designed ITS can successfully complement and substitute other instructional models at all educational levels and in many common academic subjects (Ma et al., 2014). Kulik and Fletcher (2016) in their review of Intelligent Tutoring Systems concluded that "ITSs typically raise student performance well beyond the level of conventional classes and even beyond the level achieved by students who receive instruction from other forms of computer tutoring or from human tutors" (p. 70). One of the working ingredients of these platforms is the provision of feedback. In their review of effects of feedback in computer-based learning environments, Van der Kleij et al. (2015) showed that elaborated feedback (EF; e.g., providing an explanation) produced larger effect sizes (0.49) than feedback regarding the correctness of the answer (KR; 0.05) or providing the correct answer (KCR; 0.32). EF was particularly more effective than KR and KCR for higher order learning outcomes. The larger effect sizes were found for mathematics compared with social sciences, science, and languages. Appropriate help-seeking behaviour in relation to feedback use, can positively influence student learning gain (Tai et al., 2016). Help-seeking influences students' learning when interacting with ITS as such systems often provide answers to students' requests. The role of metacognitive skills and cognitive factors is important in help-seeking behaviour (e.g. Aleven et al., 2003, 2016; Aleven, 2013; Roll et al., 2011), as the use of help can also be harmful for learning. Help-seeking in tutoring systems has been associated with students whose main objective is to finish the activity at hand and not to acquire knowledge. Although help-seeking in ITS is not a new issue, log files of student behaviour have not often been looked at in terms of sequences of activities and associated help-seeking related to performance in an online maths environment.

METHODOLOGY

Online mathematics platform

Maths-Whizz is an intelligent online tutor for 5 to 13-year-olds. It comprises more than 1200 learning objectives which have been organised into 22 topics and sequenced within each topic based on a curriculum map developed by educationalists. Maths-Whizz is being used by hundreds of schools in eight international territories and serves over 150,000 students worldwide. The platform has been used in several studies (e.g. Clark & Whetstone, 2014; Schatten et al., 2014, 2015; Mavrikis et al., 2018; McIntyre, 2022a; 2022b), but to the best of my knowledge not with an emphasis on help-seeking. An example of feedback for a 'Shape and Space' lesson is provided in Figure 1, with the feedback at the top given if an incorrect answer is given[1].

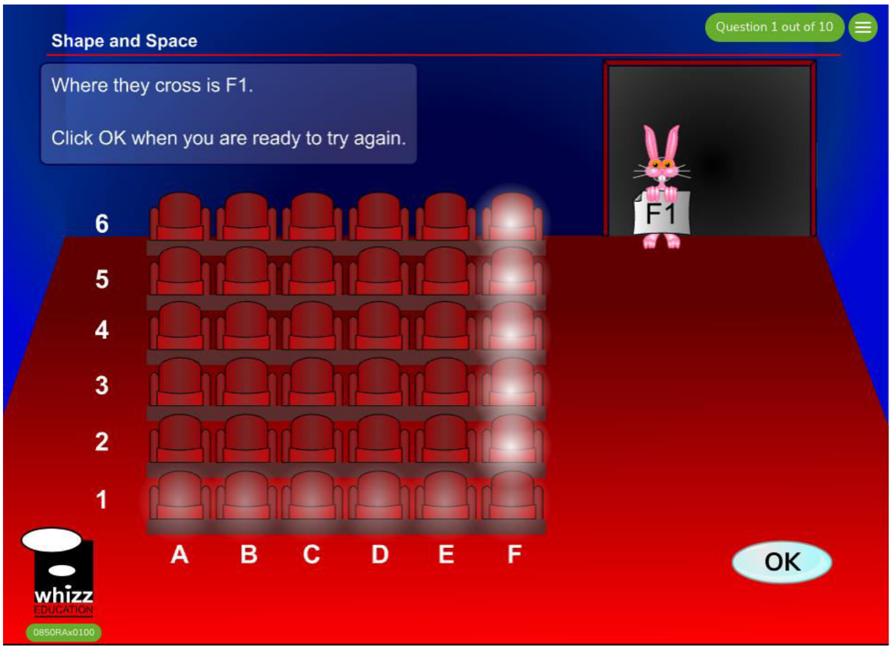

Figure 1. Feedback example for an incorrect answer in a 'Shape and Space' lesson.

Sample

The data sample included all records of school users in the UK in Years 3-5 who had at least 100 lesson records in academic year '18-'19 (N=1,799). Number of records ranged from 100 to 4970 entries, totalling 1,048,575 records from between December 2010 and January 2019.

Dataset

The dataset included 1,048,575 records, with each row corresponding to a single attempt at a lesson. Each lesson has a criterion-referenced measure of difficulty. Maths-Whizz lessons usually consist of an exercise (interactive, animated, scaffolded helps) and a test (modelled on paper tests). The dataset contains every attempt at a completed exercises and tests. Table 1 presents the variables in the dataset, as well as derived variables.

Variables	
id	A student id.
diff	The average difficulty of all exercises done by the student.
qtime	The total time the student spent completing exercises (sum of all the times).
nquestions	The total number of exercises a student completed.
totalscore	The total score for all the questions done.
totalhelp	Total help used by the student.
nlessons	Total of number of lessons (each having a number of exercises).
start	First date of Maths-Whizz usage.
end	Last date of Maths-whizz usage.
dob	Date of birth stored as integer.
gender	Categorical variable: 1=male, 2=female.
Derived variables	
range	Last date of usage minus first date of usage (end-start), in days.
precision	This was the total score divided by the number of questions (totalscore/nquestions).
intensity	This was the number of lessons per time unit (nlessons/range).

Table 1. Variables included in the dataset.

Finally, for every student there was a sequence of 'run modes'. The length of the sequence could differ per student. The seven 'run modes' could be: p=test, x=exercise, j=exercise that was jumped (i.e. student was performing well and was given a pass before completing all question, a=assessment mode (i.e. student completed the test in the diagnostic initial assessment), b=exercise that was jumped backwards (i.e. student scored 0/4 and was assigned a fail), r=replay, and s=static (i.e. student scored between 30% and 70%; it is neither a pass or a fail and means they'll receive the same lesson when returning to the topic).

Data analysis

After presenting descriptive statistics, two analyses were applied to the dataset. Firstly, multivariate regression was applied with dependent variable ‘precision’, and independent variables ‘diff’, ‘qtime’, ‘intensity’, ‘totalhelp’, ‘dob’ and ‘gender’, to see if any of these were associated with precision. Secondly, sequence analysis was applied. Sequence analysis is a data-driven approach that can unearth sequence characteristics given a dataset with sequence element. Investigating behavioural logs “has a great deal of interest” (Boroujeni & Dillenbourg, 2018, p. 206) in educational data mining and learning analytics communities. The TraMineR package, an R-package for mining, describing and visualizing sequences of states or events, is used (Gabadinho et al., 2011). In this paper we report the results of clustering to typify the types of sequences in the dataset, as well as the results of a discrepancy analysis. Such an analysis can represent the discrepancy between sequences and visualise the results in a regression tree, highlighting how the state sequences are related to one or more covariates (see Studer et al., 2011).

RESULTS

This section reports on the descriptive statistics, regression analysis and sequence analyses. Table 2 presents the descriptive statistics.

Variable	***N* = 1,799**
precision	0.87
intensity	0.79
diff	861
qtime	1.01
nquestions	2,709
totalscore	2,288
totalhelp	369
nlessons	377
range	507
dob	39,933
Gender -	Male 961 (53%) – Female 838 (47%)

Table 2. Descriptive statistics of the variables.

Table 3 presents the results of the regression analysis.

	scale(precision)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.00		
diff	0.02	-0.03 – 0.06	0.402
qtime	0.42	0.36 – 0.48	**<0.001**
intensity	0.04	-0.01 – 0.08	0.112
totalhelp	-0.63	-0.70 – -0.57	**<0.001**
dob	0.03	-0.01 – 0.07	0.171
gender	0.03	-0.01 – 0.07	0.151
Observations	1799		
R^2 / R^2 adjusted	0.191 / 0.188		

Table 3. Results of the regression analysis.

The analysis shows that both the total time the student spent completing exercises and the total help used by the student are predictive of precision, with the other variables not being significant predictors. However, there where the coefficient for total time spent is positive, that for total help is negative, indicating that more help negatively impacts on precision. Note, though, that the study does not have an experimental design, and therefore the results are *correlational*. It also is sensible to assume that students with more precision (i.e. higher achievement) need less help. Sequences from 1,799 students were included in the sequence analysis, with the minimum sequence length being 100, as per the inclusion criterion under 'Sample', and the maximum sequence length being 4,970. For determining the optimal number of clusters, we used several methods (elbow, silhouette, gap), indicating a solutions of k=3. Figure 2 presents the three cluster solutions. It is apparent that the sequence length is a confounder for the cluster solution; in other words, students differ significantly in their engagement with the learning environment. However, it can also be seen that for Type 3, the 'replay' mode, is much more prevalent, possibly indicating help-seeking behaviour. As some derived variables correlate with variables used to calculate them, not all variables were included in the discrepancy analysis, only diff, dob, gender, precision and totalhelp. If qtime we would be included, it would be a large confounder, as visually clear from the clustering. Table 4 presents the results of the discrepancy analysis. Four of these are significant variables, but based on the pseudo R-squared totalhelp contributes most to the current model, far ahead to the second most contributing variable of precision. For the regression tree, see Figure 3, we only include these two variables, as visualisation of more variables would negatively influence readability. Just like the discrepancy analysis, we do not include qtime and intensity in the regression tree, as qtime and intensity are trivially associated with the sequences of exercises done. So, we included diff, dob, gender, precision and totalhelp.

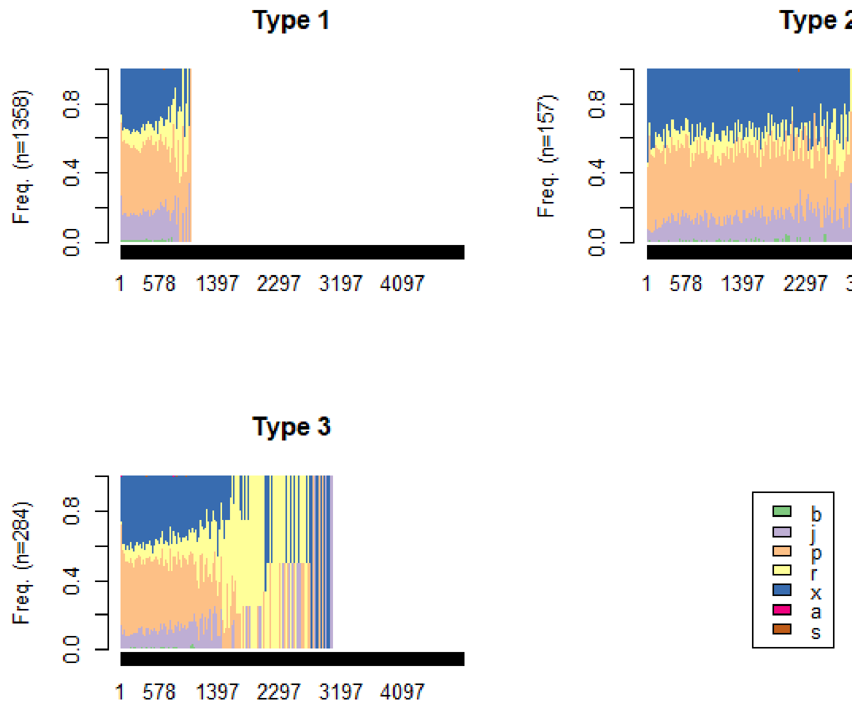

Figure 2. Three clusters of solutions.

Variable	PseudoF	PseudoR^2	p_value
diff	12.50181	0.005251	0.001
dob	18.15211	0.007625	0.001
gender	0.896066	0.000376	0.491
precision	44.77168	0.018806	0.001
totalhelp	553.9927	0.232703	0.001
Total	117.5366	0.246855	0.001

Table 4. Results of discrepancy analysis.

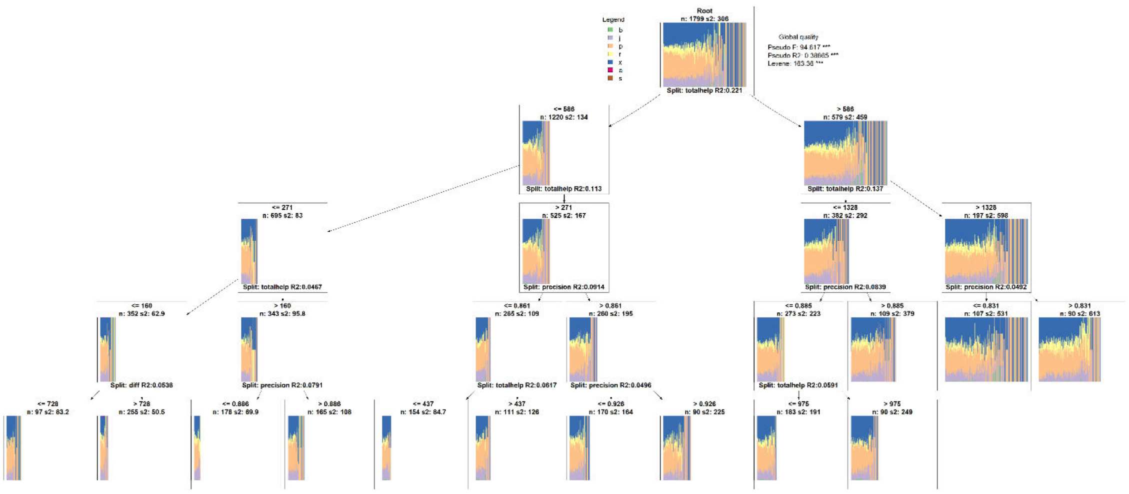

Figure 3. Regression tree for sequence analysis.

Figure 2 indicates that totalhelp and precision, as well as difficulty to a lesser extent, interact with eachother. More help extends the sequence length and extends precision, but only up to a point. Furthermore, for too difficult questions (bottom left in the diagram) help does not increase the sequence. My interpretation here is that well-tailored help helps the student along without compromising difficulty. However, this is subject to a Goldilocks 'sweet spot': help-seeking does not contribute to precision if the learning content is too difficult or the student too proficient. Help is not needed if the learning content is too easy. Help-seeking might not be productive if a student is not proficient enough yet.

CONCLUSION

This paper looked at data from an online maths environment that can be considered an ITS: students can seek for help by making use of feedback provided by the platform. The results show that feedback and associated help-seeking may not be unequivocally positive or negative. Its usefulness depends on the way it is implemented, and interacts with numerous metacognitive and cognitive factors. These results, using a learning analytics approach through sequence analysis, echo those from Aleven et al. (2016) who adjusted their initial view of the importance of on-demand, principle-based help during tutored problem solving, to a view as "helpful under certain circumstances" (p. 205).

ACKNOWLEDGMENT

Great thanks to Junaid Mubeen, the staff at Maths-Whizz, and the students who worked with the platform.

REFERENCES

Aleven, V. (2013). Help seeking and intelligent tutoring systems: Theoretical perspectives and a step towards theoretical integra-tion. In R. Azevedo and V. Aleven (Eds.) *International Handbook of Metacognition and Learning Technologies* (pp. 311–335). Springer. https://doi.org/10.1007/978-1-4419-5546-3_21

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205–223. https://doi.org/10.1007/s40593-015-0089-1

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*(3), 277–320. https://doi.org/10.3102/00346543073003277

Boroujeni, M. S., & Dillenbourg, P. (2018, March). Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 206– 215). https://doi.org/10.1145/3170358.3170388

Clark, A. K., & Whetstone, P. (2014). The impact of an online tutoring program on mathematics achievement. *The Journal of Educational Research*, *107*(6), 462–466. https://doi.org/10.1080/00220671.2013.833075

Du Boulay, B. (2016). Recent meta-reviews and meta–analyses of AIED systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 536–537. http://doi.org/10.1007/s40593-015-0060-1

Gabadinho, A., Ritschard, G., Müller, N., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37. https://doi.org/doi:10.18637/jss.v040.i04

Kulik, J.A., & Fletcher, J.D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Martinez, R., Yacef, K., Kay, J., Al-Qaraghuli, A., & Kharrufa, A. (2011). Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In *Proceeding of International Conference on Educational Data Mining 2011 (EDM 2011)* (pp. 111– 120). CEUR-WS

Mavrikis, M., Schlepps, D. S., & Mubeen, J. (2018, June). Short-term efficacy and usage recommendations for a large-scale implementation of the Math-Whizz tutor. In *International Conference on Artificial Intelligence in Education* (pp. 452–457). Springer, Cham. https://doi.org/10.1007/978-3-319-93846-2_85

McIntyre, N. A. (2022a). Access to online learning: Machine learning analysis from a social justice perspective. *Education and Information Technologies*, 1–46. https://doi.org/10.1007/s10639-022-11280-5

McIntyre, N. A. (2022b). Accelerating online learning: Machine learning insights into the importance of cumulative experience, independence, and country setting. *Computers and Education: Artificial Intelligence*, 3, 100106. https://doi.org/10.1016/j.caeai.2022.100106

Mousavinasab, E., Zarifsanaiey, N., Kalhori, S.R.N, Rakhshan, M., Keikha, L., & Saeedi, M.G. (2018). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, *29*(1), 142–163. https://doi.org/10.1080/10494820.2018.1558257

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and instruction*, *21*(2), 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Schatten, C., Janning, R., & Schmidt-Thieme, L. (2015). Integration and evaluation of a matrix factorization sequencer in large commercial ITS. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1). https://doi.org/10.1609/aaai.v29i1.9380

Schatten, C., & Schmidt-Thieme, L. (2014). Adaptive content sequencing without domain information. *CSEDU 2014: Proceedings of the 6th International Conference on Computer Supported Education*, *1*, 25–33. https://doi.org/10.5220/0004753000250033

Steenbergen-Hu, S., & Cooper, H., (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*(4), 970–987. https://doi.org/10.1037/a0032447

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331–347. https://doi.org/10.1037/a0034752

Studer, M., Ritschard, G., Gabadinho, A., & Müller, N.S. (2011). Discrepancy analysis of state sequences. *Sociological Methods and Research*, *40*(3), 471–510. https://doi.org/10.1177/0049124111415372

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, *85*(4), 475–511. https://doi.org/10.3102/0034654314564881

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

[1] This example is from https://whizz.com/year-3-maths-games/