

Evaluating the AFEL Learning Tool: Didactalia Users' Experiences with Personalized Recommendations and Interactive Visualizations

Seren Yenikent¹, Peter Holtz¹, Stefan Thalmann², Mathieu d'Aquin³ and Joachim Kimmerle¹

¹ Leibniz-Institut für Wissensmedien (IWM), Schleichstr. 6, 72076 Tübingen, Germany
{s.yenikent, p.holtz, j.kimmerle}@iwm-tuebingen.de

² Know-Center Graz; Inffeldgasse 13, 8010 Graz, Austria
sthalmann@know-center.at

³ Insight Centre for Data Analytics, National University of Ireland, Galway
mathieu.daquin@insight-centre.org

Abstract. Learning technologies offer opportunities for users to enhance and to personalize self-regulated learning activities. In this paper, we present the results of a laboratory study that covers the user evaluation of the AFEL Didactalia app, which analyzes everyday learning activities of learners, extracts learning scopes and trajectories, and provides personalized recommendations of learning resources as well as an interactive visualization of learning activities. Participants of the study ($N=76$) engaged with the tool after first completing an assigned learning task related to geography or history using learning materials from the Didactalia platform for approximately 30 minutes and afterwards freely exploring the platform for 30-45 minutes. The related behavior data was tracked and analyzed by the AFEL learning tool. After completing the two tasks, participants received an introduction to the tool and explored them as well. The results suggest a satisfactory experience with the tool and provide insights on the potential benefits, as well as aspects to be improved for further development. We discuss our findings and the next steps of the investigation in detail.

Keywords: learning analytics, mobile learning, user experience, technology-enhanced learning.

1 Introduction

This study presents the first part of a more comprehensive ongoing investigation that aims to understand the nature of everyday learning and the effectiveness of the AFEL (Analytics for Everyday Learning) tool in particular. For this purpose, we obtained feedback from users in a controlled laboratory environment. The main purpose of the project is to understand the potential benefits of tools supporting analytics for everyday learning by taking human dimension as well as design factors (i.e., content, technology) and social factors (e.g., social norms) into account [5].

1.1 AFEL Learning Tool

AFEL - Analytics for Everyday Learning is a project funded by the EU research and innovation program *Horizon 2020* that aims to device methods and tool to understand and improve informal online learning (<http://afel-project.eu>). Within the present study,

we had users evaluate the version of the AFEL tool that was specifically customized and adapted for the *Didactalia* platform (<http://didactalia.net/>), which is run by the Spanish company GNOSS, one of the partners in the AFEL consortium. Didactalia is a collection of more than 100.000 educational resources, such as web pages, slides, or interactive maps as well as educational games. Most of the materials are in Spanish language, but there is a substantial amount of English language resources as well. Didactalia also provides some social media functionalities, such as forums and groups to its users. The *AFEL Didactalia App* monitors all activities of a given learner on the platform and extracts, for example, learning scopes and trajectories [2]. In consequence, it provides personalized recommendations and an interactive visualization of the learner's activities. The main goal of the personalized recommendation function is to provide learners with user-based resources grounded in their interactions with the content. The visualization tool, on the other hand, enables personalized data exploration for the learners by offering various visualization aids, such as bar charts and plots. In the present study, we wanted to have users spend a sufficient amount of time within a controlled laboratory environment as means of generating enough data for the AFEL Didactalia App to demonstrate the aforementioned functionalities (personalized recommendation and interactive visualization) so that the learners can evaluate their usefulness.

Apart from evaluation purposes, the study was also used to relate certain forms of searching behavior to knowledge outcomes that were measured using standardized knowledge tests which had already been used in earlier studies using participants who were recruited from crowd-working platforms [3; 9]. The present study allowed for studying '(online) Search as Learning' [4] processes in a controlled laboratory environment by means of combining questionnaire data, behavioral data, and data from knowledge tests. However, the present summary of the study focuses merely on the evaluation results of the AFEL Didactalia App's recommendation and visualization functions.

1.2 Technology Acceptance Model

The theoretical foundation of the evaluation study is based on the Technology Acceptance Model [TAM; 8] that explains the process of accepting a new technology by highlighting the influence of several interrelated dimensions: external factors (e.g., system characteristics), cognitive responses (e.g., technological self-efficacy), as well as social factors (e.g., behavioral norms) lead to intermediary factors, such as perceived usefulness, perceived ease of use, and attitudes toward the technology, which in turn predict intentions to use the technology in the future. We used a specific adaptation of the TAM [7] for the e-learning context (see Method section for detailed information).

Previous research has shown the impact of user perception on improving learning analytics tools. For instance, [10] analyzed forum discussions of an online university and demonstrated that technology acceptance factors, such as intentions and community factors, played important roles in virtual academic environments. [11]

suggested that a user-centered evaluation method is necessary to understand how users *feel* when they engage with visualization systems. User perception is also essential in designing recommender applications. Apart from accuracy and quality of the content, it has been shown that usability of a system and learners' attitudinal and behavioral intentions are important features that influence systems' evaluations [12]. These studies support that approaches, which take learner needs into careful consideration and design the systems accordingly in order to offer appropriate activities and assessments for the learners, are likely to lead to deep learning [5]. Further research on evaluating learning technology in informal learning situations and across contexts requires a profound understanding of the user needs [13]. Our study extends such research that emphasizes user experiences in evaluating and designing tools for learning analytics by adopting an experimental approach based on a sound theoretical basis.

2 Method

2.1 Participants

Participants were recruited via an online participant tool that mainly consisted of university students. Participants were invited to the laboratory. In total, we had 76 English-speaking participants (54 female; average age=23.95y; median=22 y).

2.2 Material

The questionnaire for the evaluation of the experience of using Didactalia for learning tasks with the help of the AFEL Didactalia App consisted of three sections:

General evaluation. The general evaluation section was adapted from a previous study on the evaluation of software products [7], and comprised of the following dimensions that were derived from the aforementioned TAM: Perceived usefulness of the software, perceived ease of use, attitudes toward using the software, the behavioral intention, subjective norms, technological self-efficacy, and system accessibility. Each of these dimensions was measured with three items except for attitudes toward using the app (two items) and system accessibility (one item). The internal consistency (Cronbach's Alpha) of these subscales ranged from $\alpha=.69$ (subjective norms) to $\alpha=.93$ (behavioral intention). All items were answered on a seven-point Likert scale ranging from 1 = not at all to 7 = yes, definitely.

Evaluation of the recommendation function. Participants answered five evaluation questions with regard to the recommended learning resources, also using a seven point Likert scale. The items addressed their actual usage of the function (*I have checked out most of the recommended learning resources*), usefulness of the function (*The recommended learning resources were useful for me*), novelty (*The recommended learning resources were novel to me*), difficulty (*The recommended learning resources*

were too difficult for me), and the topic coverage (*The recommended learning resources were covered many different relevant areas*). Participant also answered one open question (*What did you miss in the recommendations?*).

Evaluation of the visualization function. Participants answered three Likert-type questions with regard to the visualization functionalities of the app. They were asked to indicate their usage of the function (*I have tried out the app's interactive visualization functions*), usefulness (*The visualizations were useful for me*), difficulty of the function (*It was very difficult to use the visualizations*) along with one open question (*What did you miss in the visualizations?*)

2.3 Procedure

The study took place in a laboratory with computers provided to each participant in a cubicle. At the beginning, every participant received extensive information regarding the scope and background of the study as well as data storage and data management issues and other legal aspects. Afterwards, participants provided their written informed consent and received an information sheet detailing their tasks. First, participants were assigned to either the *geography* or the *history* learning task (38 participants each). Afterwards, they filled in the respective online knowledge test for the first time. Then, participants spent 30-40 minutes looking for information regarding their learning topic on the Didactalia platform after registering to the platform using a prefabricated ID-code. After finishing their learning task, they answered the knowledge test for the second time. In the next step, participants were asked to spend 30-45 minutes freely exploring Didactalia. Finally, they were introduced by the examiner to the AFEL Didactalia App and received a 2-page leaflet explaining the app. After spending 5-10 minutes exploring the app, they answered the evaluation questions. Participants received a financial compensation of 16€ for two hours of work.

In addition to the questionnaire data, behavior data during the use of Didactalia and the mobile application was captured in order to gather information on what resources, searches and features the users have been engaged with. We also collected data in the forms of pre- and post-knowledge tests to assess the actual learning after using the learning tool. This data will be analyzed in the next steps of the investigation.

3 Results

In this section we present the findings of the general evaluation of the app as well as the evaluation of the recommendation and visualization functions.

3.1 3.1 General evaluation of the app based on TAM

We first scrutinized participants' responses to the general evaluation questionnaire based on the psychometric dimensions of the scale. In particular, easiness ($M=4.08$, $SD=1.45$) and accessibility of the app ($M=4.76$, $SD=1.76$) were among the top rated

features. Participants reported a comparatively high level of technological self-efficacy ($M=4.85$, $SD=1.37$). The ratings for perceived usefulness ($M=3.69$, $SD=1.74$), attitudes ($M=3.48$, $SD=1.47$), and subjective norms ($M=3.87$, $SD=1.37$) were also above the scale midpoint. Although these six subscales were somehow satisfactorily rated, behavioral intention to use the app in the future ($M=2.57$, $SD=1.45$) was the lowest among the seven subscales. However, according to a Kolmogorov-Smirnov test behavioral intention was neither for the history, $D(37)=0.15$, $p<.05$, nor for the geography condition, $D(38)=0.22$, $p<.001$, normally distributed. Visual inspection of the distribution of scores indicated in both cases a bimodal distribution: Whereas a majority of users did not intend to use the app in the future, there was in both cases a minority that indeed intended to use the app (see Fig. 1). Additionally, we compared the geography and the history groups in terms of the ratings they reported for each subscale. No significant differences were observed between the groups. Fig. 2 displays the mean results as well as the t and p values.

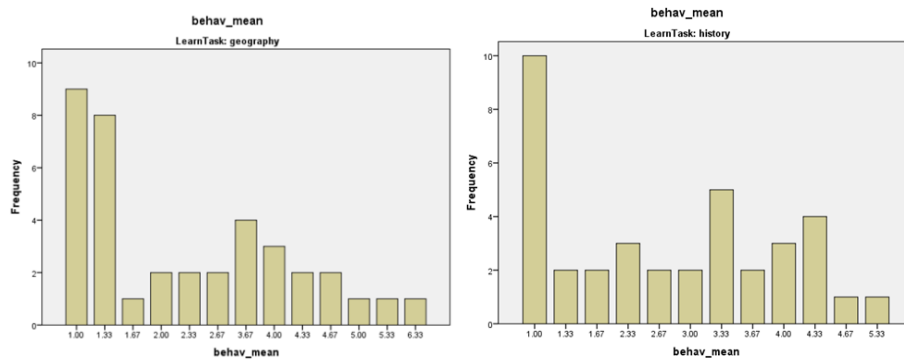


Fig. 1. Distribution of behavioral intention subscale for geography (left side) and history (right side) groups.

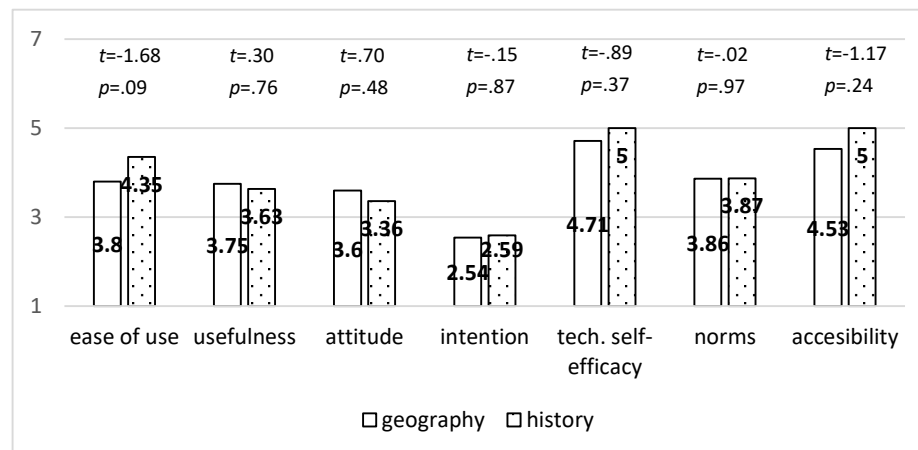


Fig. 2. Comparison of the learning tasks based on the general evaluation sub-scales. The figure provides means for each group as well as the t and p values for the comparison of the groups.

3.2 Evaluation of the recommendation function

The learners' responses to the items for the evaluation of the recommendations were in all cases slightly above the scale midpoint. Participants reported an above midscale use of the recommended sources ($M=3.79$, $SD=1.60$). The recommendation function was found moderately useful ($M=3.95$, $SD=1.55$) and novel ($M=3.91$, $SD=1.42$), and not too difficult ($M=2.45$, $SD=0.97$). Diversity of the topics covered were rated as fairly high ($M=4.15$, $SD=1.63$). In total, 45 participants answered the open question (*What did you miss in the recommendations?*). What they missed in general were English sources, more relevant sources to the assigned learning topic, and more diverse topics including other disciplines. These experiences might have kept participants from further utilization of the recommended resources and may be the reason for the moderate use of the function. Comparing the two learning topics (geography and history), we found a significant difference with regard to the use of the recommendation function, $t(71)=3.95$, $p<.001$. Learners in the history group reported more usage ($M=4.46$, $SD=1.36$) than those in the geography group ($M=3.11$, $SD=1.54$). There were no significant differences regarding the other features. Fig. 3 shows the general tendencies of the two groups.

3.3 Evaluation of the visualization function

Participants reported an above mid-point use of the visualization function ($M=4.85$, $SD=2.04$), which reflects that they were instructed explicitly to try out the interactive visualization. Although reported difficulty ($M=4.84$, $SD=1.78$) was relatively higher than in case of the recommendations, the visualization function was still rated as fairly useful ($M=4.11$, $SD=1.92$). 47 participants indicated their opinion on what they missed in the visualizations. Apart from structural features such as colors, bigger and more detailed illustrations, participants reported to have missed more diverse topics, more specific information on the searched topics and information on how to use the function. There was no significant difference between the ratings of geography group ($M=4.86$, $SD=2.17$) and the history group ($M=4.84$, $SD=1.93$) in terms of the use of the visualization function, $t(73)=0.04$, $p=.96$. The groups' responses did not differ for the rest of the features either (see Fig. 3).

4 Conclusion

Overall, the presented results provide insights on the strong aspects of the app as well as the aspects that need further improvement. In general, the evaluation of our app could be considered as satisfactory especially in terms of easiness, accessibility, usefulness, and compatibility with the subjective norms. The recommendations and the interactive visualization were also rated satisfactorily. The only statistical difference between the learning topics was observed for the learners' in the history and the geography groups in terms of the utilization of the recommendations; here, participants in the history condition indicated a higher use of recommendation sources. The lack of statistically significant differences between the two conditions with regard to other ratings of the recommendation and visualization features indicates that the app could address

different types of learning scopes. In spite of such fairly high ratings of the app features, participants were on average not very much willing to use the app in the future. This result is not in line with previous research that demonstrated effects of user satisfaction on usage behavior [6], and gets even more remarkable considering that our participants' perceived tech-savviness was quite high. It should be noted though that we found a bimodal distribution of the intention-to-use scores indicating that whereas a majority of participants is not interested in the app, a minority nevertheless intends to use it in the future. One potential reason could be related to most users' habit of using mobile apps for non-instructional purposes (e.g., texting, looking up simple information such as events). Thus, many users may not completely be aware of the benefits of such technology for learning purposes even though they acknowledge the usefulness of mobile learning apps [1].

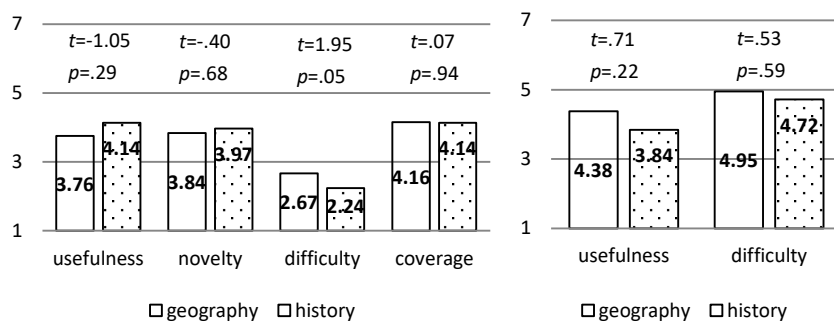


Fig. 3. Comparison of the learning groups based on the evaluations of recommendation (left side) and visualization functions (right side). The figure provides means for each learning group as well as the p - and t -values for the comparison of the groups.

The presented findings are based only on self-reports. In the next steps, we will utilize the behavioral data that was tracked during the plug-in and compare the aspects of the perceived usage and the actual usage to get a more accurate grasp of the learners' interaction with the app [6]. The results of the knowledge tests will also provide a better understanding on to what extent the AFEL app could be used to enhance learning.

Acknowledgement

The project "AFEL – Analytics for Everyday Learning" is funded under the Horizon 2020 of the European Commission (project number GA687916). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth.

References

1. Dahlstrom, E., Walker, J. D., & Dziuban, C. (2013). ECAR study of undergraduate students and information technology. 2013.
2. d'Aquin, M., Adamou, A., Dietze, S., Fetahu, B., Gadiraju, U., Hasani-Mavriqi, I., ... & Sola, S. L. (2017). AFEL: Towards measuring online activities contributions to self-directed learning. In Proceedings of the EC-TEL 2017 workshop ARTEL: Awareness and reflection technology enhanced learning.
3. Gadiraju, U., Yu, R., Dietze, S., & Holtz, P. (2018, March). Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In Proceedings of the 2018 Conference on Human Information Interaction&Retrieval (pp. 2-11). ACM.
4. Gwizdka, J., Hansen, P., Hauff, C., He, J., & Kando, N. (2016, July). Search as learning (SAL) workshop 2016. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 1249-1250). ACM.
5. Kennedy, D. M., Vogel, D. R., & Xu, T. (2004). Increasing opportunities for learning: Mobile graphing, pp. 493-502.
6. Limayem, M., & Cheung, C. M. (2008). Understanding information systems continuance: The case of Internet-based learning technologies. *Information & management*, 45(4), 227-232.
7. Park, S. Y. (2009). An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Journal of Educational Technology & Society*, 12(3), 150-162.
8. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
9. Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P., & Dietze, S. (2018). Predicting User Knowledge Gain in Informational Search Sessions. arXiv preprint arXiv:1805.00823.
10. Nistor, N., Baltas, B., Dascălu, M., Mihăilă, D., Smeaton, G., & Trăușan-Matu, Ș. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior*, 34, 339-344.
11. Cawthon, N., & Moere, A. V. (2006, July). A conceptual model for evaluating aesthetic effect within the user experience of information visualization. In Proceedings of the Information Visualization (IV'06) (pp. 374-382). IEEE.
12. Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In Proceedings of the fifth ACM conference on Recommender systems (pp. 157-164). ACM.
13. Thalmann, S.; Ley, T.; Maier, R.; Treasure-Jones, T.; Sarigianni, C.; Manhart, M. (2018) "Evaluation at Scale: An Approach to Evaluate Technology for Informal Workplace Learning Across Contexts", *International Journal of Technology Enhanced Learning* 10(4), 289-308.