Tracking and analysing the learner behaviour of non- and low-literate adults in an online literacy training environment

Eva Malessa[1]

Abstract. This study investigated what log files can reveal about learner behaviour of low- and non-literate adults learning to read for the first time in Finnish as a second language. The participants' reading development was supported by practising in an online training environment. Log files, automatically created user-computer interaction records, were chosen as empirical evidence as their analysis enables in-depth post-activity exploration of student behaviour. The quantitative analysis resulted in user profiles containing information on learner engagement, performance and productivity. Overall, the results demonstrate that individual learning performance, process, and progress can be studied and reflected on holistically by investigating the individual's digital learning footprints, their log files. Log files are an accurate and precise, yet currently underemployed research tool. More easy-to-use tools for non-experts are in demand, as current Data Mining (DM) tools are designed for computer scientists and need to be developed further to become accessible and applicable by practitioners and educational researchers.

Keywords: log files, learner behaviour, late literacy, computer-assisted language learning.

1. Introduction

A growing number of low- and non-literate adults are immigrating to highly literate countries. According to Cucchiarini, van de Craats, Deutekom, and Strik (2013, p. 96), 10-15% of the European immigrant population is estimated to be non- or

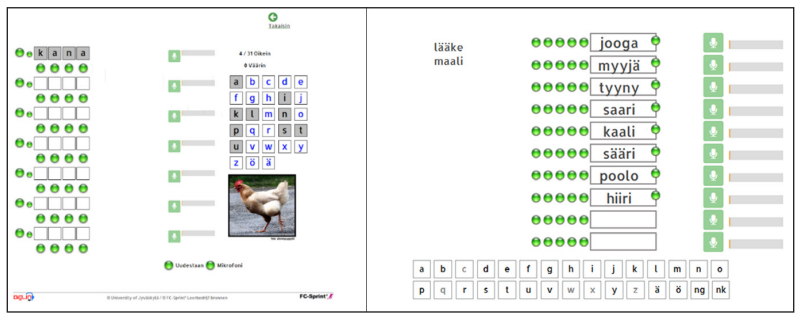
1. University of Jyväskylä, Jyväskylä, Finland; eva.i.malessa@jyu.fi

How to cite this article: Malessa, E. (2018). Tracking and analysing the learner behaviour of non- and low-literate adults in an online literacy training environment. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 182-187). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.834

low-literate. In Finland, the world's most literate country (Miller & McKenna, 2016), adult non-literacy is highly unusual and consequently, there is a paucity of academic research on how non-literate adults acquire skills in Finnish as their second language (for a review see Malessa, 2018). In practice, basic language courses are often insufficient to achieve functional literacy, even in the very transparent Finnish orthography (see Tammelin-Laine & Martin, 2015).

The 'Digital Literacy Instructor' (DigLin) was pioneered to provide Computer-Assisted Language Learning (CALL) support to low- or non-literate migrants learning to read in English, Finnish, German, or Dutch. The project (2013-2015), funded by the European Commission, enabled the development of the DigLin software which provided systematic instruction in sound-letter connections, basic decoding, and word recognition. The DigLin training environment included 300 words with seven different exercise types, including 'Listen and drag the letters' (DL) and 'Listen and drag the words' (DW), see Figure 1 below.

Figure 1. Screenshot of a DL (left) and DW (right) task



The aim of the current mixed-method study was to analyse the learners' log files automatically generated during the use of the software and to explore what log files can reveal about learner behaviour.

2. Method

2.1. Tracking computer-user interaction with log files

This study's seven participants tested the DigLin software for a period of four to six months (for details see Malessa & Filimban, 2017). The learners' software

use, including mouse/keyboard movements and microphone recordings, was automatically tracked by log files. Time-stamped log files provide detailed and objective tracking data that can be employed to make inferences about learner knowledge, processes, and strategies (Chapelle, 2007, pp. 98-99). They are currently underemployed mainly due to the extensiveness of their collected interaction records (Bruckman, 2006, p. 1449), illustrated by Figure 2.

Figure 2. Extract of a DigLin log file


```
7632;"[""04FIN""314""]";"FIN";"2014-10-30 09:21:20";"2014-10-30 09:23:58";"Drag the letters
4a";"[{""type"":""play_word_sound""  data"":""sauna""  timestamp"":""2014-10-30
09:21:20""  data_extra"":""""}  {""type"":""hide_word_picture""  data"":""""
timestamp"":""2014-10-30 09:21:21""  data_extra"":""""}
{""type"":""show_word_picture""  data"":""sauna""  timestamp"":""2014-10-30
09:21:21""  data_extra"":""""}  {""type"":""letter_drag""  data"":""s""
timestamp"":""2014-10-30 09:21:26""  data_extra"":""""}
{""type"":""letter_drag_right""  data"":""""  timestamp"":""2014-10-30 09:21:28""
data_extra"":""""}  {""type"":""letter_drag""  data"":""a""
timestamp"":""2014-10-30 09:21:32""  data_extra"":""""}
{""type"":""letter_drag_right""  data"":""""  timestamp"":""2014-10-30 09:21:34""
data_extra"":""""}  {""type"":""letter_drag""  data"":""u""
timestamp"":""2014-10-30 09:21:34""  data_extra"":""""}
{""type"":""letter_drag_right""  data"":""""  timestamp"":""2014-10-30 09:21:42""
data_extra"":""""}  {""type"":""letter_drag""  data"":""n""
timestamp"":""2014-10-30 09:21:44""  data_extra"":""""}
{""type"":""letter_drag_right""  data"":""""  timestamp"":""2014-10-30 09:21:45""
data_extra"":""""}  {""type"":""letter_drag""  data"":""a""
timestamp"":""2014-10-30 09:21:49""  data_extra"":""""}
{""type"":""letter_drag_right""  data"":""""  timestamp"":""2014-10-30 09:21:49""
```


Figure 2 presents the user interaction in a DL exercise. The documentation (workload) of dragging the letters for the word *sauna* (09:21:21-09:21:49 =29 seconds) emphasises the comprehensiveness of log file raw data. The workload contains details regarding the exact start/end date of the event, exercise type, type of actions taken, data involved, and provided feedback (right/false).

2.2. Raw data preparation and DM procedures

DigLin's log file dataset provided the empirical data for this study. Log files were accessed via phpMyAdmin and extracted for pre-processing, followed by DM, in which data is stored electronically in existing databases and the search is automated or augmented by computers (Witten, Frank, & Hall, 2011, p. 3). An initial analysis of the raw dataset supplied 3,141 log files. The data was then prepared for DM.

An event log dataset of 2,497 log files was created for the qualitative analysis, excluding no event data. For DM of this extensive database, computerised, Educational Data Mining (EDM) tools were investigated. However, faced with a limited timescale, non-expert computing skills, and the lack of an easy-to-use tool, EDM was conducted manually with the computer software Excel. Furthermore, the scope of this study's qualitative analysis was limited to 133 log files. The data were restricted to the exercise types DL and DW (see Figure 1) as both focus on the initial stages of reading development, training visual/aural grapheme-phoneme correspondences.

3. Results and discussion

This study's motivation was to investigate "what learners actually do, not what the researcher assumes instructions and task demands will lead learners to do" (Swain, 1998, p. 80). The results indicate that even though the testing sessions were relatively long (averaging 60 minutes), users were actively engaged, spending their time almost exclusively on-task. Log files' workloads record user actions and system feedback, thus providing information on how successful users perform in specific exercises. The overall success rate for letter drags in DL was unexpectedly high (78.81%), however, the users' performance was only studied for the specific skills trained in the exercises and therefore universal statements about learner proficiency are impossible to make. Further, the log files revealed that learner productivity did not equal learner performance, as the most industrious decoder was not the most successful, nor the most successful the most productive.

Qualitative log-file analysis showed that learners employed various ways to solve tasks and all strategies were not equally well-suited for all users. In many instances, the lack of successful strategies indicated an inability to learn independently, while increased autonomy and decoding proficiency were seen to stem from an increased use of efficient strategies. As log files track every single event, they show whether and how often learners make use of the provided help tools. In DL users could press buttons to listen to letter and word sounds for the words they were decoding, in DW they could listen to the words' letter sounds and were provided with an additional helptool, a soundbar (see Figure 1). The results indicate a correlation between learner proactivity, using the tools independently, and decoding success.

"Sometimes the absence of activity can be as revealing as its presence" (Bruckman, 2006, p. 1451), and log files prove that learners do not always do what they are expected to, e.g. independently exploring and employing all provided resources

(word sets, exercises, help tools). These revelations should be taken into account to enhance the CALL application's design and effectiveness. Additionally, the results also emphasise weaker students' need for more instruction and help regarding successful strategies.

4. Conclusions

This study has been a challenging, yet rewarding exploration of a new realm, log files. In sum, even though manual DM is very time-consuming and cumbersome, the results also show that it is not impossible. Nevertheless, this study acknowledges that the manual mining procedure applied to the overwhelming abundance of log files made the analysis highly prone to human error and possibly weakened the scientific rigor of the current study to some degree. Log files provide unique and innovative research data, but easy-to-use tools for non-experts are urgently needed to benefit from the valuable knowledge hidden away in the computer mines. As EDM is a relatively young research field, it remains to be seen *when* not whether EDM can make a contribution to research "in terms of providing tools and techniques that educational technology researchers can easily grasp and apply to their own research" (Angeli et al., 2017, p. 227).

References

Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: can it make a contribution? *Computers & Education, 113*, 226-242. https://doi.org/10.1016/j.compedu.2017.05.021

Bruckman, A. (2006). Analysis of log file data to understand behavior and learning in an online community. In J. Weiss, J. Nolan, J. Hunsinger & P. Trifonas (Eds), *International handbook of virtual learning environments* (pp. 1449-1465). Springer. https://doi.org/10.1007/978-1-4020-3803-7_58

Chapelle, C. A. (2007). Technology and second language acquisition. *Annual Review of Applied Linguistics*, *27*, 98-114. https://doi.org/10.1017/S0267190508070050

Cucchiarini, C., van de Craats, I., Deutekom, J., & Strik, H. (2013). The digital instructor for literacy learning. In B. Badlin, T. Hueber, G. Bailly, D. Demolin & F. Raby (Eds), *Proceedings of SLaTE 2013* (pp. 96-101). Grenoble.

Malessa, E. (2018). Learning to read for the first time as adult immigrants in Finland: reviewing pertinent research of low-literate or non-literate learners' literacy acquisition and computer-assisted literacy training. *Journal of Applied Language Studies*, *12*(1), 25-54. https://doi.org/10.17011/apples/urn.201804051932

Malessa, E., & Filimban, E. (2017). Exploring what log files can reveal about LESLLA learners' behaviour in an online CALL environment. In M. Sosiński (Ed.), *Language and literacy. Teaching LESLLA students. Proceedings of the 12th annual LESLLA symposium* (pp. 149-159). University of Granada.

Miller, J. W., & McKenna, M. C. (2016). *World literacy: how countries rank and why it matters*. Routledge. https://doi.org/10.4324/9781315693934

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge University Press.

Tammelin-Laine, T., & Martin, M. (2015). The simultaneous development of receptive skills in an orthographically transparent second language. *Writing Systems Research, 7*, 39-57. https://doi.org/10.1080/17586801.2014.943148

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018
Edited by Peppi Taalas, Juha Jalkanen, Linda Bradley, and Sylvie Thouësny

Publication date: 2018/12/08

Typeset by Research-publishing.net
Cover theme by © 2018 Antti Myöhänen (antti.myohanen@gmail.com)
Cover layout by © 2018 Raphaël Savina (raphael@savina.net)
Drawings by © 2018 Linda Saukko-Rauta (linda@redanredan.fi)

ISBN13: 978-2-490057-22-1 (Ebook, PDF, colour)
ISBN13: 978-2-490057-23-8 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-21-4 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, UK: British Library.
Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: Décembre 2018.