Doctoral School of Economics, Business and Informatics

COLLECTION OF THESES

Bálint Mátyus

Exploring the opportunities of student collaboration outside contact hours in higher education

Ph.D. dissertation

Supervisor:

László Mohácsi PhD
associate professor

Budapest, 2021

Department of Computer Science

COLLECTION OF THESES

Bálint Mátyus

Exploring the opportunities of student collaboration outside contact hours in higher education

Ph.D. dissertation

Supervisor:

László Mohácsi PhD
associate professor

Table of contents

1 Former research, justification for the topic

One of the critical components of higher education courses is the individual preparation of students. Results-oriented measuring tools are intended to reflect the achievement of educational objectives, but do not provide information on learning-effectiveness. Revealing students' learning activity and time spending habits is useful to course development process and provide new insights to better understand learning in e-learning environments. (Nagy, 2020)

The time spent on individual preparation can be measured in a difficult and inaccurate manner by means available in Hungarian higher education practice, thus it is of limited use to course development. Through student feedback systems, students could share the estimated time spent on each course. Self-reported values are likely to be biased, so it is necessary to develop methods that reveal student preparation time in an automated, objective way. The question arises: under what circumstances and conditions can student learning activity be recorded and how accurately is it possible to estimate individual study time.

Exploring the distribution of student activity over time is also essential to better understand the process of searching answers to individual questions during learning. Student procrastination narrows the window of opportunity to raise questions to tutors, and there is no time left to fill knowledge gaps. Based on my experience as a tutor over approximately a decade, I have concluded that there may be untapped opportunities for knowledge sharing among students. The idea is supported also by the works of Falchikov (2001); Kolosai *et al.* (2018); Johnson and Mays (2019). Due to the time-optimizing and procrastinating attitude of students, it is likely that the number of

students studying at the same time increases prior to exams, as well the number of questions raised. If fellow students are also involved in answering questions of their peers, the number of potential respondents may increase during more active periods. By quantifying students' learning activity, this positive aspect of procrastinating behaviour can also be revealed.

Measurement of user activity is also interesting in terms of exploring learning intensity. The question is whether students focus only on the specific task in their learning activities or share their attention among other activities.

The available methods to measure learning intensity, study time distribution and the campaign-like nature of student learning are limited. To monitor the phenomenon by objective methods, and thus to create metrics, a new e-learning system should be developed and implemented to higher educational practice.

The dissertation examines the conditions and forms of real-time synchronous question-answer (Q&A) interactions between students in an e-learning environment. The necessary, but not sufficient condition for these interactions is the simultaneous presence of participants.

The purpose of the dissertation is to create methods

1. to quantify student study time,
2. to identify periods (time-windows) outside lectures when more students study in parallel,
3. to explore student answer searching behaviour and patterns.

The following research questions are formulated:

Q1) In the examined student community, which communication platforms do students use to ask questions and what are the common answer searching patterns?

Q2) What requirements should be met by an e-learning system that supports real-time Q&A communication and allows student activity to be logged in higher education?

Q3) Using historical data provided by a learning management system, with what methods and accuracy can students' active presence be estimated?

Q4) Given the measured user activity, can it be decided for a given course when the condition of simultaneous presence for real-time Q&A interactions is met?

2 Applied methods

The dissertation is based on four interdependent studies, during which I used mixed (quantitative and qualitative) research methods.

2.1 Literature review

According to the methodological recommendations of Adams, Khan and Baeside (2007), the aim of the literature review is to explore scientific publications related to the research topic and issues of the dissertation in terms of key theoretical perspectives, frequent research methods and results obtained. During the review, the methods used in previous publications to examine the participation of students in e-learning environments are to be explored.

2.2 Exploring student answer searching habits by a questionnaire

Using a questionnaire compiled in the dissertation, the actors and used platforms of Q&A interactions are explored in the examined student community. The questionnaire can be used to identify the technologies that the students, involved to the research, used to formulate the questions, the set of potential respondents and the possible forms of communication.

Questions of the questionnaire could be included in the university-wide survey concerning the experience of distance learning conducted by the Centre for Educational Quality Enhancement and Methodology (TDTK) at Corvinus University of Budapest in the spring semester of academic year 2019/2020. (In compliance with the request of university management, students could not have been asked to participate in a separate questionnaire.)

The exploratory research highlights the relevant characteristics of students' online learning habits. Results of the survey also has a decisive role in specifying the direction of next research steps, they serve as an input for the development of e-learning systems supporting synchronous Q&A collaboration.

2.3 Development and implementation of educational materials supporting communication

To measure simultaneously the interactions of students with educational materials and with others (lecturers and peers), it is necessary to integrate communication functions to educational materials. The case study identifies the system requirements for a curriculum management system (CMS) and reveals the technological solutions. It summarises the experience of development from initial experiments to the implementation of the TalentGraf CMS at Corvinus University of Budapest. The development took almost three years, during which time various and significantly different prototypes were made. The reasons behind the changes are also discussed in detail.

The last version used at the time of submission of the dissertation supports interactive and easy-to-use creation of educational materials based on texts, instructional videos, formulas, source codes and diagrams. It also supports peer-to-peer communication in text, audio/video and screen sharing in the web browser without installing any specialized software.

2.4 Examination of the quantification of students' online learning activity

It is essential to explore the online activities of students based on objective empirical data to understand the opportunities of real-time support relationships between peers.

Due to the Hypertext Transfer Protocol (HTTP), the log files of users' activities are event-based. The range of events in a Learning Management System (LMS, e.g.: Moodle) can be recorded on server-side (page access, material download, forum activity etc). In addition to the time stamp of events, the identity of the user who caused the event and the URL of the related resource are known, but the session end time cannot be determined without a permanent network connection. Given that it is standard practice for students to download offline materials from Moodle, the possibilities for delimiting actual learning periods are limited. To define possible time-windows in e-learning virtual environment, one needs to know the beginning and end of a user's session.

Using the self-developed TalentGraf CMS, students' time spent using the system and their activity during the session can be logged. The client-side events, such as focus change on the browser window, as well as video playback events are also loggable on the server-side, thanks to the live websocket connection between clients and the server. With these options it is possible to get a much more accurate picture of the learning habits of students compared to alternative solutions, like using system logs recorded by Moodle.

Research attempts to identify active sessions and quantify student time expenditure in two ways. The first method is a self-developed

algorithm, which is referred to as an interval-based approach, the second is the DBSCAN density-based clustering procedure. The results obtained with different parameters of the two methods are compared, on which the relevant student activity can be analysed.

3 Results of the dissertation

3.1 Literature review

The literature included in the research mainly examined the log entries of systems supporting asynchronous communication using automated tools based on log files.

Regarding technology acceptance research, time of system use is operationalized using the concepts of time length, frequency, and intensity. The publications pointed out that the frequent use of e-learning systems does not guarantee better output results, but the lack of activity leads to student drop-out and poor learning outcomes.

The reviewed works, with few exceptions, did not provide detailed methodology for measuring active time. The method used for the estimation may have a significant impact on the estimated amount of time.

The opportunities of student collaboration outside contact hours can be explored in two directions. On the one hand, by introducing a self-developed system, data to be logged can be specified according to research needs. In this way some of the interactions between students linked to the curriculum can be recorded. On the other hand, activity linked to services outside the educational system can be identified

indirectly using questionnaires. The two methods should be used in combination.

3.2 Online answer searching habits of students

Based on the experience gathered, it can be concluded that in the sample there is no identifiable dominant form of communication and platform which around the collaboration is organised. Data reflect the heterogeneous habits of the analysed student community. It is important to note that it is not possible to deduce the habits of students participating in Hungarian higher education. The experience obtained is to determine the direction of the continuation of the research.

Instant messaging systems (IM) and written educational materials (WEM) stand out from the sources of information, but a large group of students have not used social media and IM services to search for answers. The reason should be researched in the future. Developing a new platform, it is worth integrating WEMs and IM functionalities, as these two types were the most popular among alternatives.

According to the answers, the respondents primarily approached their fellow students with their unique questions. This preference may also determine the choice of technology, but explanatory research is needed to justify that relationship.

Most of the student filled out the questionnaire typically do not post questions to external forums, or actively participate in Q&A discussions, and do not rely on the personal assistance of external experts, though they may be passive visitors to these websites, blogs, wikis. Online collaboration with unknown individuals is judged heterogeneously, which may also explain their passivity.

Respondents primarily ask questions to their own peers, or, if necessary, ask students from other seminars. Typically, students had a positive attitude towards the idea of online assistance between students and based on their scores, they seem to have recognised the benefits of peer-learning. However, the reliability of these responses is difficult to assess. Empirical experience needs to be gathered after the implementation of a peer learning system to produce more accurate results.

Communication was mainly in written form, but it is interesting to note that there is a strong preference for each form of communication. Written communication is the most popular, followed by audio and video communication. Screen sharing is less common than those mentioned, which is primarily used by respondents without transferring control over PC peripherals. Finding out the exact reasons may be a new direction of research, which may include lack of confidence, data security concerns, or the nature of issues raised. Students may look for answers like a waterfall, starting with the less invasive tools, then applying more and more invasive communication forms.

For a quarter of the respondents, it is not easy to find answers online, so it is worth looking for opportunities to make the self-preparation of these students more effective. However, three-quarters of them said that they could easily find the fellow student from whom they would ask for assistance. Questionnaires were filled at a time when students typically knew each other personally, as at the beginning of the semester lectures were held offline. After the implementation of Microsoft Teams, it is worth re-examining whether students can find

help easily outside the classrooms. Based on the answer, post-pandemic student expectations can be projected.

There is no dominant perception on the part of the students regarding the role of university. Most important aspects are the acquisition of factual knowledge and developing soft skills, but the importance of building social capital, developing professional mentoring relationships and entertainment is not far behind.

Questions concerning the preference between recorded materials (allowing flexible learning scheduling) and live video streams (fixed schedule) have received extremely mixed responses. It is not possible to identify a dominant student attitude. It can be assumed that students' preference in this respect differs from one course to another, the question should be asked specifically for a course.

3.3 Feasibility study, development, and implementation of TalentGraf CMS at Corvinus University of Budapest

The study confirmed the feasibility of the planned curriculum management system, which supports social interactions between users and can be integrated into Moodle. Implementation based on WebRTC and the LTI protocol successfully provided the technological background of the courses involved in the research at semesters 2019/2020/2., 2020/2021/1. and 2020/2021/2. In addition to ensuring the availability of the educational materials, the service is also used as an important measurement tool, which keeps a record of the activity of users (after the approval of students).

Based on the lecturers' experience, students have not used all the functions of the system. The number of communications established

within the system was far below expectations. In most cases, technical problems were mentioned on the message boards, which affected everyone. Individual issues may have been raised through other technologies (e.g.: instant messaging). The question arises: what is the cause for the lack of real-time communication inside the TalentGraf CMS?

It is important to note that the implementation of the latest version into university education (2019/2020/2.) approximately coincided with the distance education introduced due to COVID-19. Microsoft Teams, had been implemented in a very short time at higher education, also supports distance learning and peer-to-peer communication between students. In the winter of 2020, Microsoft Teams appears to have become a decisive and regularly accepted instrument by students and instructors, so it is worth considering the future of some of the functions of TalentGraf. It is expected that Microsoft Teams will still be present in higher education after the end of the pandemic, so it is advisable to integrate the lecturer presentation, message board and peer-to-peer communication functions with Microsoft Teams. The status markers of each task step, student activity tracking and the ranking service are still useful functions that are unique to the developed TalentGraf CMS.

As a further development option, Microsoft Teams and the ranking service can be integrated in such a way that students using the curriculum material can reach each other through weblinks that initiate communication in Teams.

3.4 Method development to explore student active sessions and time-windows

Student activity data recorded by TalentGraf CMS were collected between 20th March 2020 and 20th June 2021, involving 510 participants. The courses involved into research were Software-technology I. and II., Systems Analysis and Design, and Database Systems. In addition to join and left events to educational materials, focus change and video playback events are recorded too. Due to security and privacy reasons, more detailed data can not be obtained from the browser.

Active sessions, which should be interpreted as a time-interval, are interesting for the research, and can be associated with activities like reading materials and solving exercises. In terms of the opportunities for collaborative relationships and exploration of student activity, it is not the time of *online sessions* that provides information, but the identification of active periods. Identifying *active sessions* is not trivial based on logged events.

Using the available data, it is possible to distinguish between active and inactive status only if it is assumed that system usage generates events. An event may occur outside of actual use, which is referred to as noise in the dissertation.

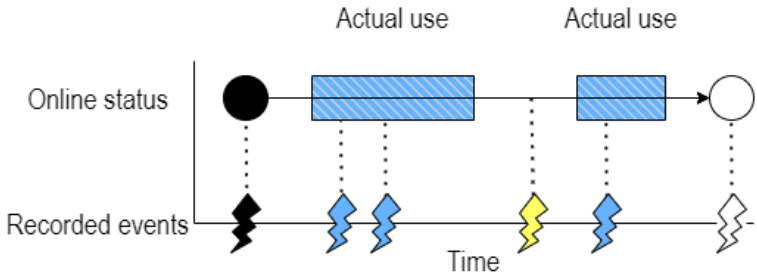

Figure 1 - Actual system use logging model
Source: own compilation

Figure 1 illustrates the relation between actual system usage and log events. The black circle indicates the beginning of a user's online session, while the white circle indicates the end of the online session. The stripped blue rectangles represent actual system usage. The logged events are displayed as lightning symbols on the time axis. The black lightning symbol indicates a join event, the white one an exit event and blue ones indicate events generated during actual usage, whereas the yellow one indicates an event generated while the system was not actively used (noise).

Active sessions of students were identified using two algorithms. The first one is a self-developed algorithm, which is referred to as interval-based algorithm, the second one is the DBSCAN density-based clustering algorithm.

None of the algorithms were more suitable than the other, according to the results obtained. DBSCAN is a widely used clustering process made popular by its tolerance to noise. DBSCAN considers the density of events too, which represents another aspect in the analysis of activity intensity.

The interval-based procedure, unlike DBSCAN, identified sessions based on join and left events. The two methods should be used in combination, which algorithm should be developed in the future.

Both methods clearly show that the more fragmented user activity is, the more inaccurate the estimation of the active time spent is. The level of fragmentation can be illustrated by the distribution of event distances, on which algorithms can be configured. Based on the statistical characteristics of the sessions generated, parameters can be estimated if the statistical characteristics converge as a function of parameter change. This was not the case for the courses examined, which draws attention to the presence of significant noise. Experience has shown that the number of clusters provided the most information: the rate of change in the number of clusters decreased gradually to larger parameters.

Input or output expectations must be set for correct parameter setting. For example, as an input expectation the analyst can specify the length of time between two events within a session that is not acceptable. As an output expectation the total time spent on the course or average session time may be mentioned. If no input or output expectation is defined, events could be clustered incorrectly.

Regarding parameter selection, it should be noted that the larger the parameters are, the longer sessions increase more than shorter sessions. It is assumed that user attention is more fragmented during longer sessions.

Empirical data do not support the assumption that students optimize time during studying. It is typical to take breaks, which could be a few minutes, or few hours long too. It is important to note that

energy optimalization can only be interpreted if all student activities are considered, including work, personal life, etc. Event and session fragmentation challenges clustering algorithms.

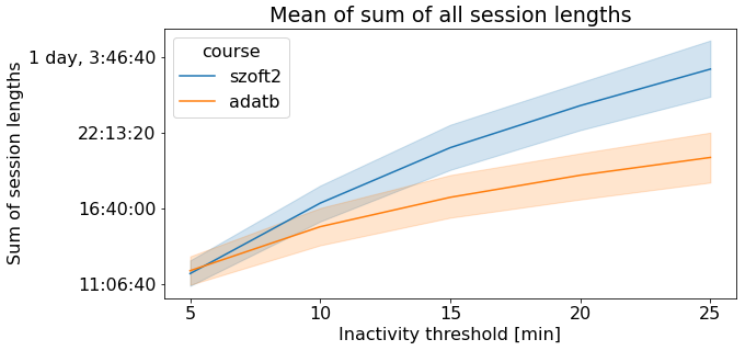

Figure 2 - Mean of sum of all session lengths based on interval-based algorithm
(t_e: 1 min) 2020/2021/2. semester
Source: own compilation

Due to the fragmentation, the sum of all session lengths of a user cannot be accurately estimated, both algorithms were very sensitive to parameters. The sensitivity analysis related to the total time spent by the 'average student' of a semester is shown in *Figure 2*. The method may be appropriate for research tasks when quantification of time expenditure is sufficient on ordinal or nominal scale. The condition for that is that the frequency distribution of event distances should not differ significantly per student.

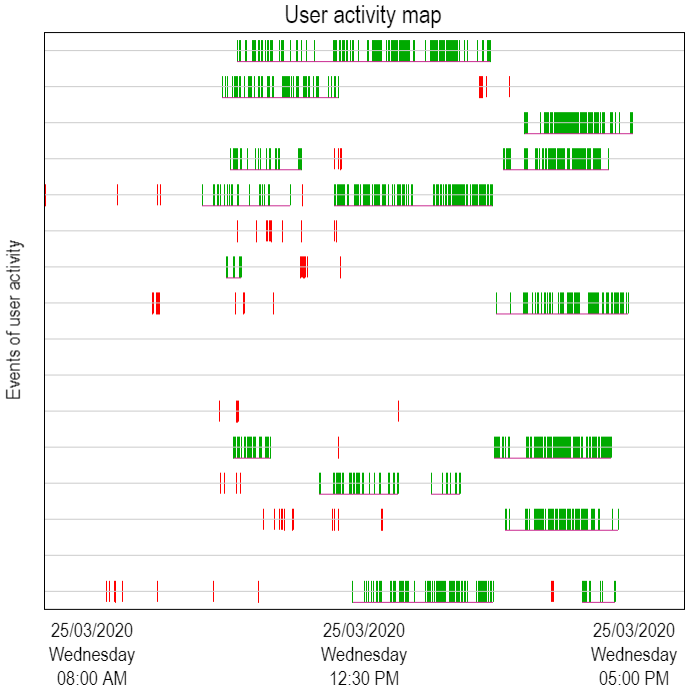

Figure 3 - Student activity map (excerpt)
25th March 2020 08:00 AM - 05:00 PM, Software-technology II.
DBSCAN, (ε = 20 p; s_{min} = 15)
Source: own compilation

Based on the activity map (*Figure 3*), students' system usage activity and the output of cluster algorithms can be seen. Each horizontal line represents a student. The red vertical lines represent the events identified as noise, while the green lines represent clustered events. The events belonging to the same cluster are connected by a purple straight line from the bottom. Based on the activity map, the parameters can be fine-tuned. There is no clear tendency to procrastination and campaign-like learning in the data. Students had heterogeneous habits in this respect.

The paper defines *time-window* as a time interval during at least two students are simultaneously active. A time-window can be considered as k ranked if there is a window of opportunity for synchronous collaboration between k users.

Most of the opportunity for the students to establish synchronous collaboration connections were in the period before examinations and on weekends. (*Figure 4*) Without lecturer intervention there are only few times when more people study in parallel. Most of the students of Database Systems were active alone in the system. It is important to note also, that after the second exam session, most of the students ‘disappeared’ from the network (as they successfully passed the test), thus the proportion of time-windows was also reduced.

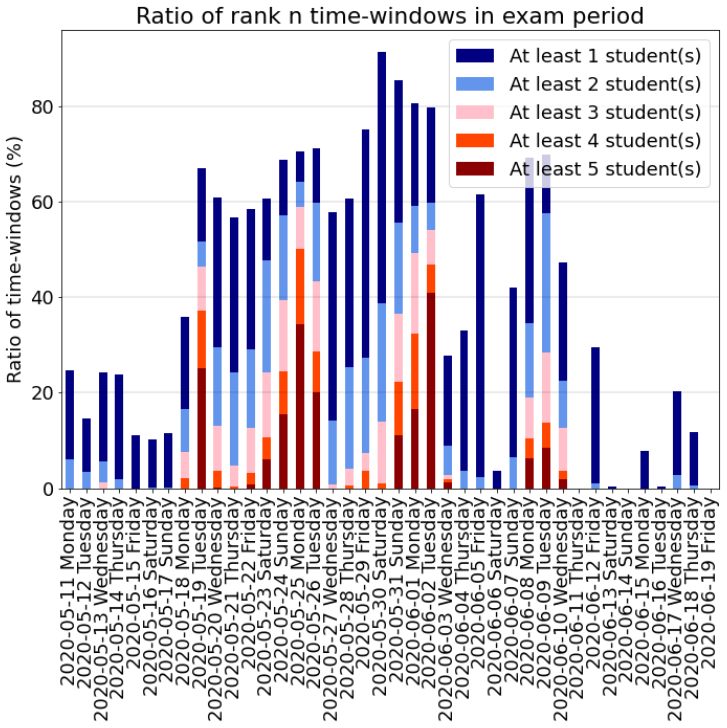

Figure 4 - Ratio of rank k time-windows in exam period based on interval-based algorithm
(t_i: 25 mins, t_e: 1 min) Database System 2019/2020/2. semester (excerpt)
Source: own compilation

In terms of quantifying time windows, the test method was sensitive to changes in parameters, but no significant differences were observed between strict and permissive parameters in terms of which days were the most likely to be used for real-time communication.

3.5 Conclusion

The dissertation examines the possibilities of real-time question-answer interactions between students working simultaneously on the same educational material.

As part of the research, a questionnaire was compiled to explore students' information searching habits and attitudes. Based on the responses, it is not possible to identify a dominant platform and form of communication. To answer the questions that arose during their self-preparation, they primarily reviewed the educational materials, and turned to their peers in the form of instant messages. These two sources of information stood out from the others. To encourage real-time collaboration relationships between examined students, it is worthwhile to direct them to the same platform. According to the results, it is justified to integrate the communication functions with educational materials.

Based on the responses to the questionnaire, the specification of requirements of TalentGraf CMS was defined, which enabled logging students' online learning activity. A feasibility study was conducted, several prototypes were developed over 3 years. The latest version was implemented after approval to Corvinus University of Budapest in the semester 2019/2020/2. The CMS system was used to observe four courses in three semesters at the University.

With the help of the collected data, it became possible to examine the time spent by students, the fragmentation of their active presence, their campaign-like learning, procrastinating behaviour, and the proportion of periods when several people used the educational materials simultaneously. The dissertation presents an analysis method

to evaluate log data. Source code is available in the form of Jupyter Notebook in the appendix.

4 Main references

Adams, J., Khan, H. T. A. és Baeside, R. (2007) Research Methods for Graduate Business and Social Science Students, Research Methods for Graduate Business and Social Science Students. doi: 10.4135/9788132108498.

Bókay, A., Derényi, A. és Temesi, J. (2016) „Az ECTS szerepe a képzési programok tervezésében és a minőségfejlesztésben", in Tanulásszervezés, tudáselismerés és átjárhatóság a felsőoktatásban. Kaposi József, o. 37–48.

Davis, F. D. (1989) „Perceived usefulness, perceived ease of use, and user acceptance of information technology", MIS Quarterly: Management Information Systems, 13(3), o. 319–339. doi: 10.2307/249008.

Falchikov, N. (2001) *Learning Together: Peer Tutoring in Higher Education.* 1. kiad. Routledge. Elérhető: https://www.routledge.com/Learning-Together-Peer-Tutoring-in-Higher-Education/Falchikov/p/book/9780415182614 (Elérés: 2021. február 19.).

Hrastinski, S. (2008) „What is online learner participation? A literature review", Computers and Education. Pergamon, 51(4), o. 1755–1765. doi: 10.1016/j.compedu.2008.05.005.

Johnson, S. N. és Mays, K. A. (2019) „A Retrospective Analysis of a Cross-Year Peer Tutoring Program for Oral Health Students", *Journal of Dental Education*. Wiley, 83(2), o. 137–143. doi: 10.21815/jde.019.016.

Kolosai, N. *és mtsai.* (2018) „A kortársoktatás elméleti és gyakorlati aspektusai – A »Tanulj, tanítsd, tudd!« program tapasztalatai", *Új Pedagógiai Szemle*. Elérhető: https://folyoiratok.oh.gov.hu/uj-pedagogiai-szemle/a-kortarsoktatas-elmeleti-es-gyakorlati-aspektusai-a-tanulj-tanitsd-tudd (Elérés: 2021. február 19.).

Mihályi, Z. (2019) „A felhasználói élmény tervezés és a motiváció kapcsolata egy webes projektben”.

Nagy, V. (2020) *E-learning értékelési módszerek*.

Venkatesh, V. *és mtsai.* (2008) „Predicting different conceptualizations of system USE: The competing roles of behavioral intention, facilitating conditions, and behavioral expectation”, *MIS Quarterly: Management Information Systems*. Management Information Systems Research Center, 32(3), o. 483–502. doi: 10.2307/25148853.

5 Own publications regarding the topic

5.1 Studies published in referenced professional journals in foreign language

Molnár T., Kő A., Mátyus B. (2017) Exploring Usability and Acceptance Factors of m-Government Systems for Elderly. In: Kő A., Francesconi E. (eds) Electronic Government and the Information Systems Perspective. EGOVIS 2017. Lecture Notes in Computer Science, vol 10441. Springer, Cham

Mátyus B. (2021) Monitoring the duration of use of web-based systems by cluster-based procedures. In: SEFBIS Journal, HU ISSN 1788-2265 (megjelenés alatt)

5.2 Studies published in scientific books, book chapters, peer-reviewed conference proceedings in foreign language

Kő, A., Molnár, T. and Mátyus, B. (2018). A User-centred Design Approach for Mobile-Government Systems for the Elderly. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/8631531. doi: 10.1109/SKIMA.2018.8631531.

5.3 Other studies

Mátyus, B. (2017) Exploring Usability and Acceptance Factors of m-Government Systems for the Elderly [Powerpoint Presentation] Research Forum, 30 May, 2017, Corvinus University of Budapest. Available at: http://web.uni-corvinus.hu/~y7qbqz/ppt/kutforum.pdf

Mátyus B. (2018) TalentGraf CMS prototypes [Computer program]. Available at: http://web.uni-corvinus.hu/~y7qbqz/

Mátyus B. (2021) Online learning activity visualization instrument [Computer program]. Available at: http://web.uni-corvinus.hu/~y7qbqz/