e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 4 No. 2, Desember, 2020, **Hal. 164-171**
DOI: 10.29408/edumatic.v4i2.2688

Evaluation of Rapidminer-Aplication in Data Mining Learning using PeRSIVA Model

Muhammad Zamroni Uska[1], Rasyid Hardi Wirasasmita[2], Usuluddin[3], Baiq Desi Dwi Arianti[4]
[1,2,4] Department of Informatics Education, Universitas Hamzanwadi
[3] Department of English Education, Universitas Hamzanwadi
email: zamroniuska@gmail.com[1], rasyidhw.pinformatika@hamzanwadi.ac.id[2], usuludin15@gmail.com[3], ariantibaiq@hamzanwadi.ac.id[4]

(Received: 30 Oktober 2020 / Accepted: 30 November 2020/ Published Online: 20 Desember)

Abstract

RapidMiner is an application that is used to analyze data quantities and qualitatively to obtain information and knowledge as expected. This software is implemented to process data using several methods or algorithms in Data Minig learning. However, when using this software, users sometimes cannot distinguish between various methods or algorithms in Data Mining. Therefore, it is necessary to evaluate to optimize the use of this software in data mining learning. This study focuses on RapidMiner evaluation of data mining learning using the Persiva model. This model consists of aspects of satisfaction, behavior, impact, and effectiveness. The data collection technique was in the form of a questionnaire with 48 subjects. Data analysis used is descriptive statistics to determine satisfaction, behavior and effects. Meanwhile, Think-Aloud Retrospective technique is used to determine the effectiveness of RapidMiner. Our findings show that users are satisfied with the results of respondents on average agreeing (80%), in the aspect of behavior and impact, the percentage results are above 80%, and the use of this application has been effective with a completion rate above 90%. So it can be concluded that by using this application in data mining learning users can easily complete tasks, and be motivated, and add insights and knowledge in relevant disciplines.

Keywords: RapidMiner, Data Mining, PeRSIVA Model.

INTRODUCTION

The Data Mining course is a compulsory subject in the Department of Informatics Education, Faculty of Mathematics and Sciences (MIPA), *Universitas Hamzanwadi*. The application of this course is related to data analysis (data management), the results of which can obtain knowledge or can determine patterns in an event. Data Mining is the process of extracting data or data processing that can be used to obtain knowledge (Han, Pei, & Kamber, 2011; Shu, Sliva, Wang, Tang, & Liu, 2017). In addition, addition, data mining can also be interpreted as exploration or Knowledge Discovery in Database (KDD) (Parsaei, Rostami, & Javidan, 2016; Ristoski & Paulheim, 2016; Siddique, Akhtar, Khan, & Kim, 2019). Each of the term's knowledge and knowledge discovery has a concrete definition all relationships in the context of data mining.

In learning data mining at *Universitas Hamzanwadi*, it discusses how to find patterns or knowledge by using several methods or algorithms, some of the main roles of data mining, namely: estimation, prediction, classification, clustering, and association. In Estimation and Prediction (using Linear Regression and Neural Network methods), classification (using the Naive Bayes method, K-Nearest Neighbor and C4.5), clustering (using the K-Means method and Self-Organizing Map (SOM), and Association (FP-Growth and A Priori) In Data Mining learning process. This software used as a media or learning tool is Rapid Miner, where RapidMiner makes it easier for students (user) to data process or analyze very large amounts of data (Çelik & Başarır, 2017; Vyas & Uma, 2018). However, in the learning process, some students in inputting data and processing with the help of RapidMiner, could not distinguish between the use of the naïve Bayes and K-Nearest Neighbor methods in data classification, clustering using the k-means and SOM methods, and so that errors occur in the RapidMiner output. In addition, their ability to analyze data is low. This is due to the understanding of concepts and capabilities in using the RapidMiner application on different students.

Based on the foregoing, a form of program evaluation or learning media is needed that can reveal all the abilities / potentials possessed by students, especially those related to the competency standards of this subject. Efforts to make better academic quality can be pursued through improving the quality of learning and the quality of the media used (Littenberg-Tobias & Reich, 2020) (Sunder M, 2016). The best learning system will produce a good learning quality as well, and this shows the relationship between these two aspects. In this case, the intended evaluation is RapidMiner-Aplication evaluation of Data Mining learning, which is used as a tool for data analysis to find patterns or information that can be used as information and knowledge.

The evaluation aspect cannot be ignored by a lecturer or teacher. Evaluation is a continuous process that forms the basis of all good learning activities. In general, evaluation can be defined as a systematic process to determine the value of something (objectives, activities, decisions, performance, processes, people, or objects) based on certain criteria. The purpose of evaluation is to obtain accurate and objective information about a program or media that has been implemented. This information can be in the form of program implementation process, impact / results achieved efficiency. The results of the evaluation can also be used as a benchmark for whether the media or program is successful or not, can be continued or stopped, and can be used as a basis for developing follow-up programs.

Several evaluation models that have been developed and are very popular and widely used as strategies or work guidelines in the implementation of evaluation of learning programs include: 1) Kirkpatrick's Four Levels Evaluation Model, 2) Evaluation of the CIPP Model (Context, Input, Process, and Product), 3) Evaluation of the Stake Model (Model Maintenance), 4) and the PeRSIVA model. Each of these four models has different characteristics. In this study, it is more focused on the evaluation of the PeRSIVA model.

The PeRSIVA model is a model developed by (Chrysafiadi & Virvou, 2013b), where this model is a combination of the Kirkpatrick model which consists of four levels (reaction, learning, behavior, and results) and the evaluation framework Brusilovsky, Karagiannidis, & Sampson consists of two aspects between them: user modeling and decision making. The PeRSIVA model consists of four aspects of assessment, namely: satisfaction, behavior, impact, and student model validity (effectiveness). Satisfaction is one of the important criteria to justify the success of a system that has been developed or implemented (Chrysafiadi & Virvou, 2013b), while behavior is something that can affect the characteristics of users of the system both internal and external (Mokshin, Mokshin, & Sharnin, 2019). Furthermore, the impact (effect) aspect is said to be one of the main factors in measuring the success of the program or system that has been implemented (Tarhini, Hone, Liu, & Tarhini, 2017). Meanwhile, effectiveness is defined as the performance of a product or system (software) that has been applied (Chrysafiadi & Virvou, 2013b; Llorens, Noé, Colomer, & Alcañiz, 2015; Sorgenfrei & Smolnik, 2016)

A number of researchers have implemented some learning evaluation models or programs in various fields, including (Uska, Wirasasmita, & Fahrurrozi, 2020) (Uska, Usuluddin, & Sriatni, 2019) (Gaudioso, Montero, & Hernandez-Del-Olmo, 2012). Meanwhile, those who evaluate e-learning or the education system and adoption several different approaches include (Albano, 2011; Chrysafiadi & Virvou, 2013b; Gaudioso et al., 2012). For now, (Chrysafiadi & Virvou, 2013b) have developed this model (PeRSIVA) in programming language learning assisted by this learning -learning media. Furthermore

(Chrysafiadi & Virvou, 2013a) apply it to evaluate e-learning in learning the C programming language.

Based on this, this study applies this model which aims to evaluate Rapid Miner in data mining learning, so that it can provide an overview of satisfaction, attitudes, effectiveness, and impact on the use of software that has been implemented.

METHOD

This type of research is quantitative using survey methods. The model used to evaluate the RapidMiner in Data Mining-learning is PeRSIVA which consists of four aspects, namely: satisfaction, behavior, impact, and effectiveness. The number of subjects in this study was 48 people who had data mining learning. The data collection technique used was a questionnaire with five Likert scale options. Meanwhile, the data analysis technique used is descriptive quantitative on the aspects of satisfaction, behavior, and effects. While the Retrospective Think-Aloud technique is applied to analyze the effectiveness of the implemented of RapidMiner or media (system) that has been implemented by looking at the number of errors and the Compliance Rate in the application or system (Alhadreti, Elbabour, & Mayhew, 2017; Muntinga & Taylor, 2018; Salmerón, Naumann, García, & Fajardo, 2017).

RESULT AND DISCUSSION

Result

Learning or the lecture process is considered successful if students are satisfied with the media or learning tools that have been applied (used). In this case, learning uses the Rapid Miner media as a medium for analyzing or processing data on the Data Mining methods or algorithms. The learning process is better if they give a positive reaction to the learning-environment.

Table 1. Survey Result of *Behaviour Aspect*

Indicator	**Mean Score**	**N**
Positive Effect	3.94	48
Interest	4.13	
Motivation	3.98	
Understand	4.09	

The results of user satisfaction in the application of Rapid Miner as a tool or software in analyzing data using several data mining algorithms can be seen in Figure 1. These results indicate: 89.90% of respondents agree that Rapid Miner is easy to use; 84.09% agree it can help solve calculation problems; 78.27% agree it can simplify the analysis (calculation); 75.89 agree can increase knowledge; 43.44% agree this software is a waste of time in its use; 75.55% agree can increase motivation. Meanwhile, 82.89% agree that Rapid Miner lives up to expectations; 75.20% agree this is always used; 83.78% agree this is flexible in its use; and 87.78% agree it can speed up data analysis. Based on our findings, overall from the indicators of the satisfaction aspect, it can be concluded that the use of Rapid Miner as software in Data Mining courses is that students feel satisfied in.

The results regarding user behavior in implementing RapidMiner software are shown in table 1. These results show that the user's attitude in using this software has a positive impact or effect with an average score of 3.94, fosters student interest with a score of 4.13, can motivate with a value of 3.98, and can understand needs of students of 4.09. Based on the results shown in table 1, the results of the respondents' perception on average agreed that this software had a positive effect of 75.50%; can also generate interest by 84.20%, and can motivate by 79.90%, in addition, can understand the needs of students by 80.60% (see figure 2).

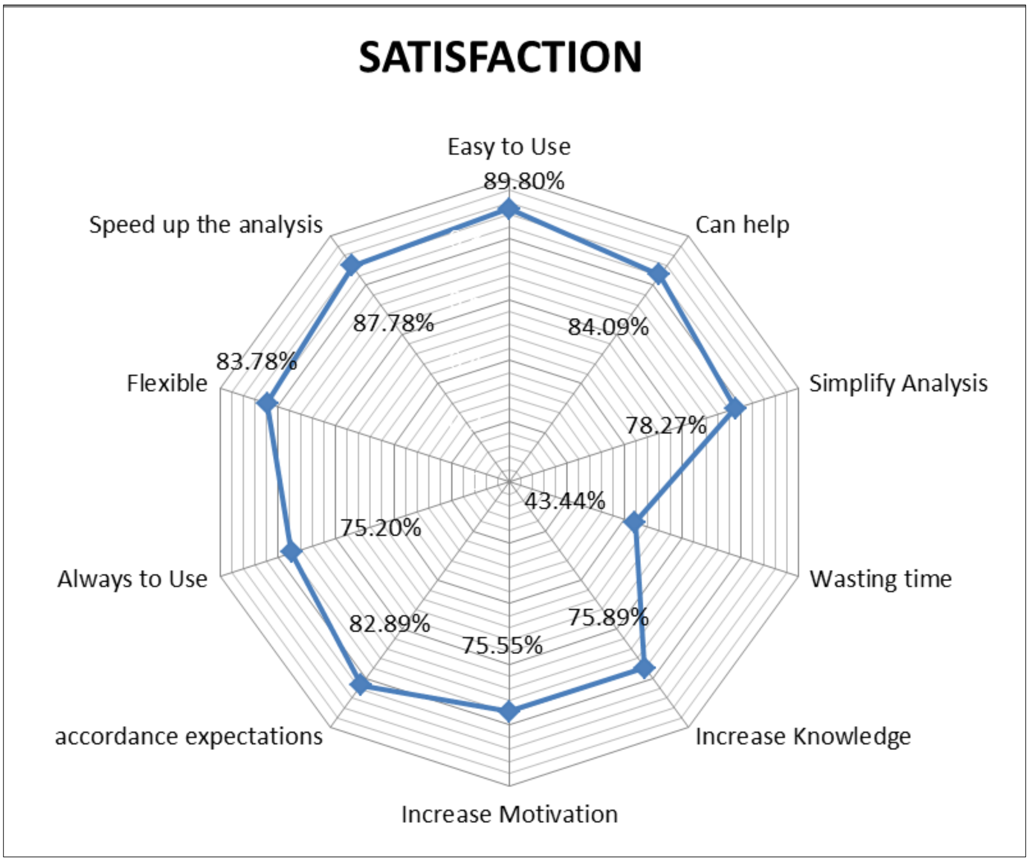

Figure 1. The Result of User Satisfaction

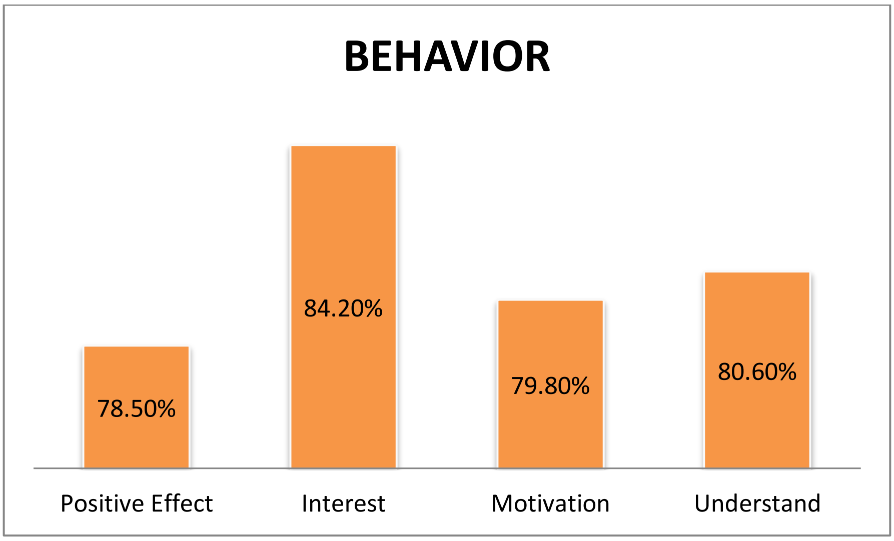

Figure 2. Percentage Results on Behavior Aspect

The results of the survey toward impact aspects in learning progress is shown in table 2, which manifest that the use of RapidMiner can understand the concept and analysis of algorithms in data mining, obtaining an average score of 4.24, can help in learning various analysis methods with a value of 3.80. In addition, this software can assist in learning (lectures) with a value of 3.98, data analysis other than data mining (4.09), and help in doing assignments and learning activities (mean score 4.15). These results indicate that with the help of this software, it can affect the learning process in accordance with what is expected.

Table 2. Survey Result of Impact (Effect) Aspect

Indikator	Mean Score	N
Understand the concept and analysis of algorithms (Data Mining)	4.24	48
Assist in learning various analytical methods on data mining	3.80	
Assist in learning (study)	3.98	
Assist in data analysis other than data mining	4.09	
Assist in doing assignments and learning activities	4.15	

In addition to the results of the table, based on the results of the questionnaire shown in figure 3, 86.07% of respondents on average agree that this software can understand the concepts and analysis of algorithms in data mining, 81.20% agree to help in learning various methods of analysis, 83.89% stated that they agreed to help in the learning process (study). Meanwhile, 84.06% agreed that they could perform data analysis other than Data Mining (DM), and could help with tasks and activities during DM lectures by 85.90%.

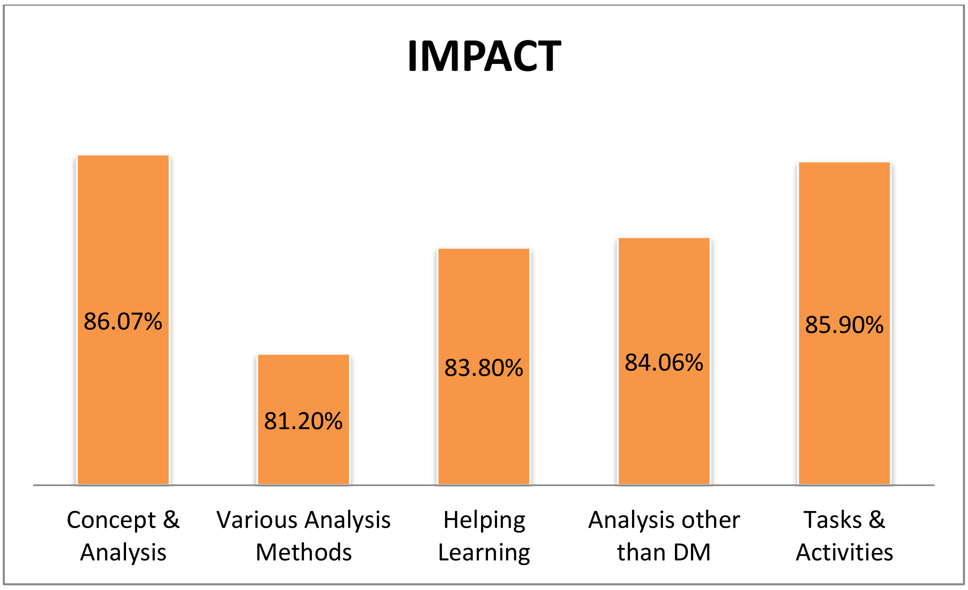

Figure 3. Percentage Results on Impact (Effect) Aspect

The results of the survey on the effectiveness of using Rapid Miner as a tool in the data mining lecture process can be seen in table 3. These results indicate that there are six types of tasks performed by users (students) related to the algorithms used to analyze data with Novice group (24 people) and an expert group (24 people). In task 1 (linear regression), they did a task regarding linear regression. Where these results are the percentage of errors when using this software is 0% with a percentage 100% Completion Rate (novice & expert).

In task 2 (naïve bayes) and task 3 (K-Nearest Neighbor), the error percentage in the novice group was 8% and the expert group was 4%. Meanwhile, the percentage of completion rate in the novice group was 92%, and the expert group was 96%. Furthermore, task 4 (K-Means) has an error percentage of 4% (novice) and 0% (expert) with a completion rate of 96% (novice) and 100% (expert). The results on task 5 (FP-Growth) had an error percentage of 4% in the novice and expert (groups) with a completion rate of 92% (novice & expert). Meanwhile, the last task is task 6 (A Priori) which has an error percentage of 8% in the novice group, while 4% in the expert group with a completion rate of 92% (novice) and 96% (expert).

Table 3. Recapitulation Results of RapidMiner Effectiveness

Task	**Number and Percentage of errors**		**Completion Rate (%)**	
	Novice	Expert	Novice	Expert
Task 1	0 (0%)	0 (0%)	100	100
Task 2	2 (8%)	1 (4%)	92	96
Task 3	2 (8%)	1 (4%)	92	96
Task 4	1 (4%)	0 (0%)	96	100
Task 5	2 (8%)	2 (8%)	92	92
Task 6	2 (8%)	1 (4%)	92	96

Discussion

The results of user satisfaction with the use of RapidMiner as a learning medium to analyze data in data mining learning get positive results. Based on the results of user satisfaction shown in Figure 1, an average of over 70 percent agreed. This indicates that the software (RapidMiner) is easy to use, can help users solve data calculation and analysis problems, and can increase their knowledge. In addition, this software is flexible, speeds up time in doing analysis, and can increase user motivation, so that the always to use this software, because it is in accordance with user expectations. These findings are in accordance with or relevant to research conducted by (Chrysafiadi & Virvou, 2013b, 2013a). Where in their research, on average, they agree that the use of e-learning and e-training learning can improve programming logic, knowledge, benefit, and can increase exploration.

Our findings on the behavior aspect, the average score obtained in this aspect are 4.04 and the number of percentages who agree on all indicators is above 75%. So these results suggest that RapidMiner has a positive effect on users, can increase user interest, increase motivation, and can increase their understanding of data analysis and calculations in data mining. Our findings are relevant to research conducted by (Chrysafiadi & Virvou, 2013b, 2013a), where their findings show that using e-learning and e-training can increase their interest and motivation, their understanding of computer learning or computer programming in language C.

Furthermore, our findings on the impact aspect are shown in Table 2. These results show that the mean score obtained for all indicators is above 3.80, and the percentage that states agree on this aspect is an average of 81%. This indicates that by implementing or using this RapidMiner as a learning medium, users can understand the concepts and analysis of algorithms in data mining and can help students learn various methods or analysis algorithms in data mining. In addition, with this software, it can help students (users) in the learning process and can assist them in analyzing data in addition to existing methods in data mining (for example: statistical analysis in general), because data mining is also related to data mining and statistical analysis. Meanwhile, this software can help users complete their tasks and activities during this lecture (data mining).

Our findings on the impact aspect are relevant to the results of research conducted by (Chrysafiadi & Virvou, 2013b, 2013a). Their findings indicate that the existence of e-learning and e-training can improve their performance in understanding computer programming (language C). Meanwhile, our findings are also relevant to the results of research conducted by (Abu-Al-Aish & Love, 2013; Adel, 2017; Yilmaz, 2017). Their research shows that e-learning and m-learning have a positive impact or effect on users in utilizing or implementing the education software (e-learning and m-learning).

Furthermore, the last one is our findings on the effectiveness aspect of using RapidMiner is by testing the capabilities of this software by performing 6 tasks (tasks) consisting of algorithms or analysis methods used in data mining, namely; linear regression, naïve bayes, K-Nearest Neighbor, K-Means, FP-Growth, and A Priori. In this test, 48 respondents used to do each of these tasks with an expert group of 24 people, and a 24 novice group consisting of a group of users who were still unfamiliar with this software.

Our findings show that in task 1 (linear regression) the completion rate is 100% in the novice group and 100% in the expert group. This means that there are no errors or errors in doing data analysis on the linear regression method. In task 2 (naïve Bayes), 2 novice users in doing this task an error occurred, while in the expert group only 1 person made an error. This indicates that in this task, participants or users do not pay attention to the explanation or understand the steps to carry out the analysis using the naïve Bayes method. In task 3 (K-Nearest Neighbor), two, participants made mistakes in the novice group, and one person in the expert group. This indicates that when they upload data, they do not pay attention to the column or row that will be uploaded, so an error occurs in the data output process, as well as what happened in task 4 –task 6, so that errors were made when using or do the task with the method or algorithm.

Based on the foregoing, even though there were errors or errors, overall the percentage of errors made by participants was very small, so that the results on the completion rate obtained a percentage above 90%. So it can be concluded that the use of RapidMiner as a tool or media in analyzing data in data mining courses is effective. Our findings are relevant to the results of research conducted by (Sinclair, Kable, Levett-Jones, & Booth, 2016; Sorgenfrei & Smolnik, 2016; Uska et al., 2020) e-learning or systems have implications for learning processes and learning outcomes, and has a positive impact on its users.

CONCLUSION

Based on the results and discussion that have been previously described, it can be concluded that the results of the PeRSIVA evaluation model in Data Mining learning, obtain satisfaction results as expected, while the results on the Bevavior aspect can have a positive effect and can increase interest, motivation, and student understanding. Meanwhile, using PeRSIVA users can understand concepts, various methods of analysis, and can assist in carrying out tasks and activities on Data Mining learning. Furthermore, this software (RapidMiner) is effectively used during the learning.

REFERENCE

Abu-Al-Aish, A., & Love, S. (2013). Factors influencing students' acceptance of m-learning: An investigation in higher education. *International Review of Research in Open and Distributed Learning*, *14*(5), 82–107.

Adel, R. (2017). Manage perceived e-learning quality in Egyptian context. *Total Quality Management & Business Excellence*, *28*(5–6), 600–613. https://doi.org/10.1080/14783363.2015.1103174

Albano, G. (2011). Knowledge, Skills, Competencies: A Model for Mathematics E-Learning. *Communications in Computer and Information Science*, 214–225. https://doi.org/10.1007/978-3-642-22383-9_18

Alhadreti, O., Elbabour, F., & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: is there added value? *Journal of Usability Studies*, *12*(3), 95–110.

Chrysafiadi, K., & Virvou, M. (2013a). Dynamically personalized e-training in computer programming and the language C. *IEEE Transactions on Education*, *56*(4), 385–392.

Chrysafiadi, K., & Virvou, M. (2013b). PeRSIVA: An empirical evaluation method of a student model of an intelligent e-learning environment for computer programming. *Computers & Education*, *68*, 322–333.

Gaudioso, E., Montero, M., & Hernandez-Del-Olmo, F. (2012). Supporting teachers in adaptive educational systems through predictive models: A proof of concept. *Expert Systems with Applications*, *39*(1), 621–625.

Littenberg-Tobias, J., & Reich, J. (2020). Evaluating access, quality, and equity in online learning: A case study of a MOOC-based blended professional degree program. *The Internet and Higher Education*, *47*, 100759.

Lloréns, R., Noé, E., Colomer, C., & Alcañiz, M. (2015). Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*, *96*(3), 418-425.e2. https://doi.org/10.1016/j.apmr.2014.10.019

Mokshin, A. V, Mokshin, V. V, & Sharnin, L. M. (2019). Adaptive genetic algorithms used to analyze behavior of complex system. *Communications in Nonlinear Science and Numerical Simulation*, *71*, 174–186.

Muntinga, T., & Taylor, G. (2018). Information-seeking strategies in medicine queries: a clinical eye-tracking study with gaze-cued retrospective think-aloud protocol. *International Journal of Human–Computer Interaction*, *34*(6), 506–518.

Salmerón, L., Naumann, J., García, V., & Fajardo, I. (2017). Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning*, *33*(3), 222–233.

Sinclair, P. M., Kable, A., Levett-Jones, T., & Booth, D. (2016). The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: a systematic review. *International Journal of Nursing Studies*, *57*, 70–81.

Sorgenfrei, C., & Smolnik, S. (2016). The effectiveness of e-learning systems: A review of the empirical literature on learner control. *Decision Sciences Journal of Innovative Education*, *14*(2), 154–184.

Sunder M, V. (2016). Constructs of quality in higher education services. *International Journal of Productivity and Performance Management*, *65*(8), 1091–1111.

Tarhini, A., Hone, K., Liu, X., & Tarhini, T. (2017). Examining the moderating effect of individual-level cultural values on users' acceptance of E-learning in developing countries: a structural equation modeling of an extended technology acceptance model. *Interactive Learning Environments*, *25*(3), 306–328. https://doi.org/10.1080/10494820.2015.1122635

Uska, M. Z., Usuluddin, U., & Sriatni, S. (2019). Usability Testing Method: Efisiensi, Efektivitas dan Kepuasan Pengguna Terhadap Website SMKN 1 Praya. *Edumatic: Jurnal Pendidikan Informatika*, *3*(2), 91–98.

Uska, M. Z., Wirasasmita, R. H., & Fahrurrozi, M. (2020). The application of Usability Testing Method for Evaluating the New Student Acceptance (NSA) System. *Journal of Physics: Conference Series*, *1539*(1), 1–6. IOP Publishing.

Yilmaz, R. (2017). Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom. *Computers in Human Behavior*, *70*, 251–260.