An Online Controlled Experiment Design to Support the Transformation of Digital Learning towards Adaptive Learning Platforms

Nathalie Rzepka[1], Katharina Simbeck[1], Hans-Georg Müller[2] and Niels Pinkwart[3]
[1]*University of Applied Sciences Berlin, Treskowallee 8, 10318 Berlin, Germany*
[2]*Department of German Studies, University of Potsdam, Am neuen Palais 10, 14469 Potsdam, Germany*
[3]*Department of Computer Science, Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany*

Keywords: Online Controlled Experiments, Adaptive Learning Systems, Digital Learning Environment, Learning System Architecture.

Abstract: Digital learning platforms are more and more used in blended classroom scenarios in Germany. However, as learning processes are different among students, adaptive learning platforms can offer personalized learning, e.g. by individual feedback and corrections, task sequencing, or recommendations. As digital learning platforms are already used in classroom settings, we propose the transformation of these platforms into adaptive learning environments. To measure the effectiveness and improvements achieved through the adaptions an online-controlled experiment design is created. Our result is a process that consists of the target definition, development of the prediction model, definition of the adaptions, building the experiment architecture, the experimental period, and the hypothesis testing. As an example, we apply this design exemplarily to an online learning platform for German spelling and grammar. In this way, we contribute to the research field by bridging the gap between adaptive learning technology and the process of transformations and experiment designs.

1 INTRODUCTION

Digital learning platforms offer many opportunities to optimize learning processes. Since the COVID-19 pandemic, they have also increasingly become part of the educational landscape and are used as a supporting medium in school lessons. A great relief for teachers, for example, is the automatic correction and the subsequent immediate feedback for users. However, traditional digital learning platforms are not yet personalized - even though it is known that students learn differently. Adaptive learning platforms offer personalized adaptations to provide each user with the optimal learning platform. This can be expressed through personalized feedback, appropriate difficulty levels, recommendations, or other individualized interventions. Since there are already existing digital learning platforms, it makes sense to consider how these can be transformed into adaptive learning platforms. It is further useful to investigate how effective the built-in adaptations are and by how much they improve the learning success.

For this aim, we propose an online controlled experiment design to transform a traditional learning platform into an adaptive learning platform. We present this design exemplarily on the platform orthografietrainer.net, a platform which is used in German school lessons and serves the acquisition of spelling and grammar competences.

We proceed as follows. First, we describe online controlled experiments and A/B-testing. After that, we introduce adaptive learning platforms and their architecture. In section 3, we define the phases of an online controlled experiment on how an adaptive learning platform can be designed, implemented, and evaluated from a digital learning environment. We exemplify each phase with the transformation of the orthografietrainer.net platform. After that, we discuss and classify our results.

2 BACKGROUND

2.1 Online Controlled Experiments and A/B Testing

Learning analytics in education often defines interventions based on machine learning (ML) based

Rzepka, N., Simbeck, K., Müller, H. and Pinkwart, N.
An Online Controlled Experiment Design to Support the Transformation of Digital Learning towards Adaptive Learning Platforms.
DOI: 10.5220/0010984000003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 2*, pages 139-146
ISBN: 978-989-758-562-3; ISSN: 2184-5026

prediction outcomes. However, these interventions need to be reviewed for effectiveness. Controlled experiments are the best scientific design, to ensure the causal relationship between intervention and changes in user behavior (Kohavi et al. 2007). A/B-tests are used on large sites such as amazon, google or Bing to calculate the effect of user interface (UI) changes of apps and websites, algorithms, or other adjustments (Kohavi and Longbotham 2017). In a simple case, A/B-tests consist of control (default version, A) and treatment (changed version, B). Users are randomly assigned to one of these versions and their actions on the website or app are logged. A previously defined overall evaluation criterion (OEC) is a quantitative measure of the change's objective. At the end, after the experimentation period, a hypothesis test is done to find out if the difference in OEC between the two variants is statistically significant (Kohavi et al. 2007). This enables data-driven decision-making in web-facing industries.

2.1.1 Randomization Unit

The randomization or experimentation unit is the item on which observations are made, in most cases this is the user (Kohavi et al. 2007). The users are randomly assigned to one variant, but the assignment is persistent. Further, the entities should be distributed equally, which means in the case of an A/B-test that the users are split up by 50%. While the distribution should be equally, best practice is to have a treatment ramp-up before (Kohavi et al. 2007). This starts with a lower percentage for the treatment which is gradually increased. Each phase runs for few hours which offers the opportunity to check for problems and errors, before it is shown to a wide range of users.

There are different designs of randomised trial, for example student-level random assignment, teacher-level random assignment or school-level random assignment. When choosing the randomisation design, there are theoretical and practical reasons, which are summarised by Wijekumar et al. (2012). Choosing the teacher-level or school-level has the advantage that users in a school class belong to the same control group. This is particularly useful if the experiment is carried out in class - otherwise the teacher would have to divide the class. Statistical power plays a role in the decision between teacher-level and school-level, as the analysis of which has shown that within-school random assignments are more efficient than school-leveled random assignments (Campbell et al. 2004). The disadvantage of choosing teacher-level or school-level assignments is a reduction in effective sample size, considering that observations withinside the cluster have a tendency to be correlated (Campbell et al. 2004).

2.1.2 Overall Evaluation Criterion

The OEC defines the goal of the experiment and must be defined in advance (Kohavi et al. 2007). It can also be referred to as a response variable, dependent variable, or evaluation metric. The definition of the OEC is of rather great importance, as the rejection of the null hypothesis is based on the comparison between the OEC of the two variants. As the experimentation period is in most cases only few weeks, the OEC must be measurable in the short-term while being predictive in the long-term. Deng and Shi differentiate between three types of metrics that can be used as OECs (Deng and Shi 2016): business report driven metrics, simple heuristic based metrics and user behavior-driven metrics. Business report driven metrics are based on long-term goals and are associated with the business performance, such as revenue per user (Deng and Shi 2016). Simple heuristic-based metrics are describing the interaction of the user on the website, for example, an activity counter. User behavior-driven metrics are based on a behavior model, for example for satisfaction or frustration. Whatever type of metric is chosen in the end, there are two important characteristics for metrics: directionality and sensitivity (Deng and Shi 2016). Directionality describes that the interpretation of the metric must have a clear direction, for example, the bigger the OEC the better and vice versa. Sensitivity means that the metric should be sensitive for the changes made in the variant (Deng and Shi 2016).

2.1.3 Architecture

There are three important architecture components of A/B-tests: randomization algorithm, assignment method and data path (Kohavi and Longbotham 2017). The randomization algorithm is the function that maps the user persistently to one variant. As stated above, the distribution between the variants should be equal. In the second step, the assignment method allocates the user to the mapped variant. This can be either by redirecting the user to a new webpage, by traffic splitting, or client-sided by dynamically adjusting the web page according to the variant changes (Kohavi and Longbotham 2017). The data path describes the component which collects the user interaction (e.g., the clickstream data) and aggregates and processes it afterwards. Another tool which should be built-in is a diagnostics system, which graphs the numbers of randomization units in each variant, metric means and further effect to inform researchers about the progress during the experimentation period (Kohavi and Longbotham 2017).

2.1.4 Hypothesis Testing

The analysis of an A/B-test is straightforward statistics with hypothesis testing. The null hypothesis (H_0) states that the OECs for the variants are not different. The treatment is accepted as being significant if H_0 is rejected. The confidence level should be 95%, which means that there is type 1 error in 5% of the cases. Although the power is not measured separately, it should be checked to be between 80%-95%. The standard error should be small and can be decreased by increasing the sample size and lower the variability of the OEC. The variability of the OEC can be reduced by triggering: it is often the case, that the tested component is not entered by all users. For instance, if the variant is implemented in the purchase process of a website, but not all users who are logging in are purchasing something. These users need to be excluded from the sample to reduce variability.

2.1.5 Limitations

There are some limitations which need to be considered when implementing A/B-tests (Kohavi et al. 2007). While the OEC can be a data-driven basis to either reject or accept the null hypothesis, it does not explain why the hypothesis should be accepted or rejected. Furthermore, an effect can only be measured during the experimentation period. The period should be chosen carefully, as effects are not registered if the period is too short. If the webpage also has experienced users, there is an effect of newness, as the users must get used to the changes first (in case they are assigned to the variant). As the variant is mostly a prototype, it should be considered that errors in the prototype effect the OEC massively. Also, performance issues of the variant are known to impact the OEC (Kohavi et al. 2007). This issue can be faced by A/A-tests, where the randomization algorithms and assignment methods are tested before implementing the treatment variant (Kohavi et al. 2007; Kohavi and Longbotham 2017). In education there are also ethical concerns: if users are assigned to a variant that works poorer, they are treated differently than the others which is unfair. Users can also find out differences on the app or webpage if they compare it to what is shown to other users.

2.2 Adaptive Learning Environments

Adaptive learning environments offer individualized learning by adjusting to its users and their learning process. Paramythis and Loidl-Reisinger (2003) define learning environments as adaptive if they are capable of "monitoring the activities of its users, interpreting these on the basis of domain-specific models; inferring user requirements and preferences out of the interpreted activities, appropriately representing these in associated models; and, finally, acting upon the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process". There are different categories of adaptive learning environments: adaptive interaction, adaptive course delivery, content discovery and assembly, and adaptive collaboration support (Paramythis and Loidl-Reisinger 2003). While adaptive interaction offers different options on the system's interface such as font size, color schemes, or restructured interactive tasks, adaptive course delivery fits the course content to user characteristics. Content discovery and assembly focuses on providing suitable learning material from distributed repositories. Adaptive collaboration support is meant to support communication processes between multiple persons. An adaptive learning environment typically consists of three components: domain model, learner model, and the tutoring model, sometimes referred to as adaptive model (Paramythis and Loidl-Reisinger 2003; Meier 2019). The domain model describes the learning content and their relationship to one another. It should represent the course being offered and involves all information about the learning objects. The learner model, also user model, contains all information about the learner, to be able to support the adaptation of the system (Brusilovsky and Millán 2007). Here, a feature-based approach is most common, less popular are stereotype-based techniques of user modeling. Brusilovsky and Millán propose the user's knowledge, interest, goals, background, individual traits, or context of work as the most important features (Brusilovsky and Millán 2007). The tutoring or adaptive model describes which and when adaptations are being offered, for example in terms of learning paths or recommendations.

When designing an adaptive learning environment, there are four approaches that can be distinguished from each other: macro-adaptive approach, aptitude-treatment, micro-adaptive approach, and constructivist-collaborative approach (Beldagli and Adiguzel 2010). These approaches describe in which way and on which basis the platform or environment is adapted. In macro-adaptive approaches, the student's profile is considered, for example prior exercises, intellectual abilities, or cognitive and learning. Here the presentation of content or language of presentations are adapted. The aptitude-treatment approaches offer different types of instructions or different types of media for different students. The micro-adaptive approach is based on on-task measurements. Here the users' behavior is monitored during the learning process and based on the stored information, the instructional design is adapted. It can be divided in two phases: the diagnostic process, during which learner characteristics and aptitude is assessed and the second phase, the prescriptive process, where the content is adapted, for example by task sequencing. The constructivist-collaborative approach includes collaborative technologies and focuses on the learning and sharing knowledge with others.

Both adaptive learning environments and online controlled experiments have already been researched in different contexts. However, to support the implementation of adaptive learning environments, the further development of an existing learning platform towards an adaptive platform can be useful. In this development process, different interventions can be tested for their effectiveness in order to find the best solution. We therefore link the two research fields of adaptive learning and online controlled experiments in our work to bridge the gap between technology and process and to propose a systematic approach to further development and evaluation.

3 TRANSFORMATION AND EVALUATION TOWARDS AN ADAPTIVE LEARNING ENVIRONMENT

The aim of this paper is to introduce an experiment design to systematically conceptualize, implement and evaluate the transformation of a digital learning platform into an adaptive learning platform using the orthografietrainer.net environment as an example. The individual phases are described both in general terms and exemplarily for the online platform orthografietrainer.net.

3.1 Orthografietrainer.net

The online platform orthografietrainer.net has existed since 2011 and contains exercises on various areas of spelling and grammar. These include, among others, capitalization, comma placement, separated and combined spelling, as well as sounds and letters. So far, the platform has been used by more than 1 million users who have completed a total of 10,4 million exercises. Most of the users are students who are registered on the platform as part of their school lessons and receive exercises as homework. The advantage for teachers is that the exercises are automatically corrected, and error corrections are directly displayed to the students. Furthermore, the platform offers evaluations for teachers so that they can quickly get an overview. The platform is therefore primarily used to accompany lessons, with grading and the teaching of the subject matter continuing to take place in face-to-face lessons.

Teachers can select tasks depending on the competence area, for example "Capitalization of time indications as adverbs and nouns". Each of the tasks consist of 10 sentences on the selected spelling problem. A special feature of the platform is the dynamic task process: if a mistake is made while working on the 10 sentences, the task expands automatically by adding more sentences that convey exactly the same thing as the incorrect sentence (a different version of the sentence). This forces a user to repeat the problem, which was obviously not understood, more often. While dynamic adaptation offers the advantage that weak points receive special focus, it can also increase frustration if sentences are repeatedly added that the student is unable to successfully complete the assignment.

For each user, the registration process provides demographic data, such as gender, region, and state, as well as grade level and type of school. Furthermore, there is learning process data, since the actions that a user performs on the platform are stored. Thus, it is possible to see which exercises the user has done and when, and what mistakes he or she has made. Further data available are details about the exercises, for example the exact solutions and the task difficulty.

The platform offers a high didactic potential, as it already includes the immediate feedback, the automatic correction and evaluation via dashboards, as well as the dynamic adjustment of the task structure (i.e., repetitions in case of wrong answers). However, the experience (task sequencing, feedback) is the same for all users, although some tasks might be more difficult for some users than for others. Thus, one approach to personalization is task selection and order.

3.2 Description of the Phases

The process of transformation and evaluation through A/B-testing consists of six phases:

(1) Target definition
(2) Development of the prediction model
(3) Definition of the adaptions
(4) Building the experiment architecture
(5) Experimental period
(6) Hypothesis testing

3.2.1 Target Definition

The analysis of the platform is at the beginning of the process and serves to get an overview of the platform, its users and usage. The number and type of users must be determined, as well as the average count of users per day. It is also important to find out in which context a platform is used (e.g., in the context of school lessons, at university and as an exercise platform shown to children by their parents). Furthermore, it should be assessed what data is stored when the platform is used and what data is already available. This could be behavioral data, demographic data, or data about the tasks (e.g., their difficulty) in addition to clickstream data. Finally, it must be clarified which implementation options are available for adaptive models.

For the transformation of a digital learning platform and the evaluation through A/B-testing, the randomization unit and the OEC should be defined. In many cases, the randomization unit is the user, however, other entities are possible too. Furthermore, it should be discussed if the randomization is student-leveled, teacher-leveld or school-leveled.

There are various metrics that can be defined as OEC, for example by the number of correctly solved tasks, the ratio of correctly solved tasks, the number of tasks solved (as a measure of stamina), or interaction with the platform (opening tips or explanations). Another option is to use competency models such as the Rasch model to calculate a competency for each person and measure how quickly and how far it has changed.

In the example of orthografietrainer.net, the users are the randomization unit. Randomisation takes place at the student-level. Since the user does the tasks at home as homework in the typical scenario, the group differences do not influence the school lessons. At the same time, the student-level randomisation assignment prevents intracluster correlation, which would reduce the effective sample size (Campbell et al. 2004).

The goal of the adaptations on the online learning platform orthografietrainer.net is to improve the aptitude of the students, which can be assessed by implementing the Rasch model (Boone 2020). The Rasch model belongs to the Item Response Theory models. Besides the analysis of competencies, it can also be used for surveys or assessments (Khine 2020). Instead of the Rasch model, one could also simply count the number of correct exercises per user. However, using the Rasch model has the advantage that the model includes the difficulty of the task, and maps task difficulty and person competence on the same scale (Boone 2020). The adaptive system thus calculates the person competence for each user using the Rasch model and continuously updates the value as the user solves new tasks. Thus, the competence of individuals in the intervention group should have increased more than that of the control group after the experimental period:

$$R'_{treatment} > R'_{control} \tag{1}$$

3.2.2 Development of the Prediction Model

Following the target formulation, the prediction model must be defined in more detail. For this purpose, it is determined which variable y is to be predicted. Furthermore, feature engineering and feature selection are used to determine which data in the model are used for prediction. Finally, several ML algorithms are tested and evaluated to find out the best one for the use case. The implementation of prediction models in the education domain are described in more detail by Xing and Du (2019) or Dalipi et al. (2018).

In this example, the probability that the next exercise will be answered correctly is to be predicted. This model is trained with existing data that has been stored over the last few years. The data includes demographic data, learning process data, and data about the upcoming task, as described above.

3.2.3 Definition of the Adaptions

The results of the prediction model are then used to offer suitable adaptions. A first step is to choose the adaption category defined by Paramythis and Loidl-Reisinger (2003), explained in section 2. After that, interventions must be defined and then be determined which interventions will be applied to which predictions. Interventions are classified by Wong and Li (2018) into four different categories: direct message, actionable feedback, categorization of students, and course redesign. Depending on the prediction model and learning platform, different interventions can be considered. It is important here to consult with educational designers and pedagogues to define interventions in a pedagogically sound way.

In the example of orthografietrainer.net, interventions based on the solution probability for the next set of exercises are defined. These are of the type adaptive course delivery, as the courses' content and presentation are adjusted. There are different types of interventions that can be tested in the experiment:

A describes the status quo, no intervention takes place. B summarizes the interventions that show the user the result of the prediction in different ways. A distinction is made between a verbal display, which translates the solution probability into a statement, and the percentage display. C describe pedagogical interventions that are used when solution probabilities are low. For example, showing the rule that must be used to solve the task or display on an example sentence. D is an intervention of the task order. Here, the order of the sentence is adjusted and only sentences with a certain probability of being solved are displayed. This is to maintain motivation as most sentences are solved successfully.

3.2.4 Building the Experiment Architecture

Once the goals have been formulated, the predictive model and interventions are developed, the experiment architecture needs to be prepared. Here the approaches of Beldagli and Adiguzel (2010) can be used (section 2). After that, the implementation of a randomization algorithm, of the assignment method and the data path follows. Depending on the number of interventions n the randomization algorithm divides the randomization unit into n groups and the assignment method maps the result of the randomization algorithm to one variant. Furthermore, it should be implemented that every interaction of the platform which is needed to calculate the OEC is stored in a database. Before the adaptions are implemented, an A/A-test should also be carried out to check the experiment setup. In the end, the experimental period is defined in whole weeks to avoid differences between weekdays and weekends.

Regarding the example of the orthografietrainer.net platform, we implement a randomization algorithm that uses the user ID to map the user to one of the variants. The assignment of users to the variant is done by redirecting them at the beginning of a session to the mapped variant. Implementing the data path includes storing every interaction of the user to be able to calculate the OECs later. We set the experimental period at eight weeks. In 2021, an average of 11,000 people per day were active on the platform. In 2019, before the pandemic, the average was 1,000 people per day. If we assume a decrease to 5,000 people per day in 2022 (because all classes take place back at school and there are no school closures due to the pandemic), eight weeks still gives us 280,000 sessions to evaluate.

3.2.5 Experimental Period

In this phase, the A/B-test is carried out. It starts with a treatment ramp-up as described in section 2. During the experimental period, the process is observed by a diagnostic system to continuously check the experiment.

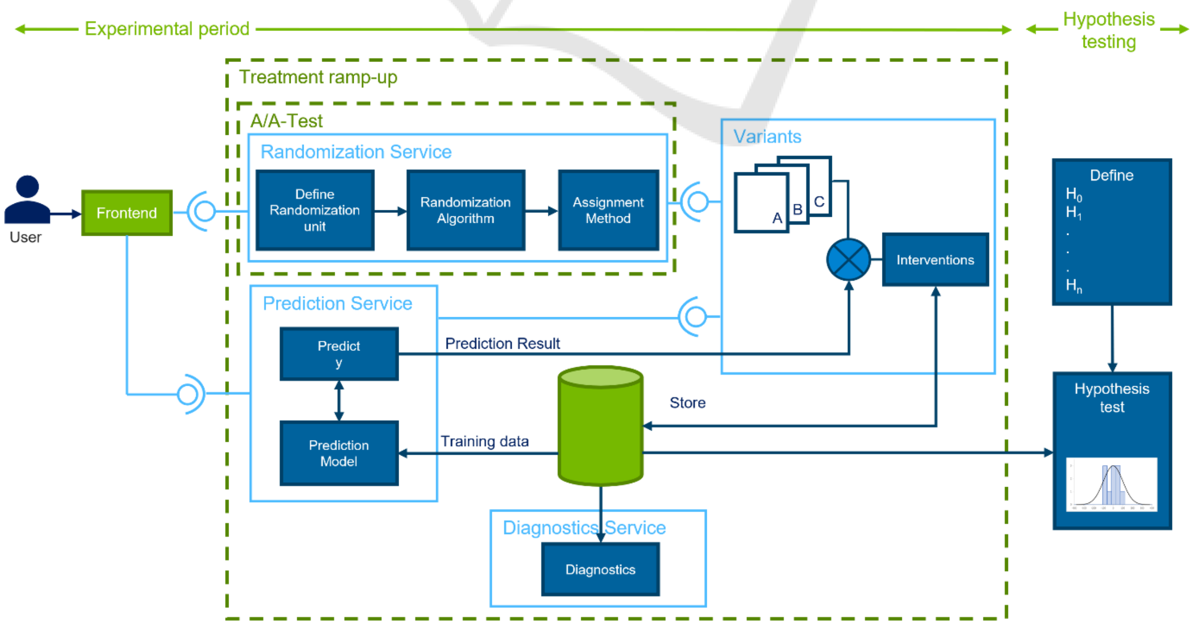

Figure 1: Experimental Architecture.

Figure 1 shows the architectural design of the online controlled experiment. To avoid changing the source code of the platform too much, the new components are implemented as services. The user first calls up the frontend as usual. There are then two interfaces, one to the randomization service and one to the prediction service. The randomization service first identifies the user as a randomization unit by means of the user ID. The randomization algorithm then maps the entity to the variant. The assignment method forwards the entity to the appropriate implementation. This randomization service is tested within the framework of A/A tests. The prediction service consists of a prediction model and the activity "predict y". The prediction model was trained and tested in advance with data from the database. The model is used to predict the user's solution probability. Both services described above have an interface to the variants. Depending on the prediction result and variant, the user is shown a suitable intervention. All user interactions with the platform, including the results of the practice sets, are stored in the database. The database is also used by a diagnostics service that monitors the equal distribution of users during the test. The whole process starts with a treatment ramp-up and then leads to an equally distributed A/B-test.

3.2.6 Hypothesis Testing

After the experimental period the OEC is calculated for each variant and the hypothesis test is carried out. It is defined by:

H_0: OEC does not differ between the variants

H_1: OEC differs between variants

$$H_o : OEC_A = OEC_X \; against \\ H_1 : OEC_A \neq OEC_X, \\ where \; X = \{B, C, D, \ldots n\} \tag{2}$$

Depending on the actual n at the end of the experimental period the test statistics are described and executed. The result of the hypothesis test leads to either accept or reject H_0.

4 CONCLUSION & OUTLOOK

We have proposed a design for an online controlled experiment that supports the transformation and evaluation of a learning platform into an adaptive learning platform.

For this purpose, the process phases presented in section 3 were run through as an example using the orthografietrainer.net platform. The adaptive learning environment presented uses predictive models and subsequent interventions to individualize the user's learning process. The development is evaluated through an online controlled experiment (A/B-testing) and subsequent hypothesis tests to examine the effect of the interventions. The next steps are the exact implementation of the defined predictive model and interventions according to Figure 1, and the measurement of effectiveness afterwards through hypothesis tests.

Our work encourages the redesign of learning platforms towards adaptive learning environments instead of developing them from scratch. In this way, transfer to real-world applications becomes easier and more practical. In addition, evaluating different interventions as part of the transformation process provides the opportunity for data-driven decision-making when implementing adaptations in learning environments.

There are several limitations for the transferability to other applications and for the execution of the experiment. One limitation is the imprecise runtime of the experiment. The decision about the runtime of the A/B test often depends on external factors. The longer the A/B test runs, the more likely it is that the long-term goals of improving competence can be measured. This also depends on the competency being measured: the variability of spelling competencies is very slow, where, on the other hand, there are rapid progresses in competency in other fields. Here, an exchange with pedagogues and educational designers is appropriate to determine an adequate duration for the experiment.

The number of interventions must also always consider how many users are expected for the test. When implementing in smaller settings than orthografietrainer.net, the number of interventions may need to be adjusted to have enough users per intervention. Furthermore, prediction models are only possible as an implementation if data are already available to train them accordingly in advance. We also recommend that the models and interventions in the experiment be additionally validated for fairness to ensure that automated decisions are not detrimental to subgroups.

Further research is planned to specifically address the implementation of the experiment and prediction model as services, so that existing platforms can be extended without having to deal with a legacy codebase.

REFERENCES

Beldagli, B.; Adiguzel, T. (2010). Illustrating an ideal adaptive e-learning: A conceptual framework. In Procedia - Social and Behavioral Sciences 2 (2), pp.5755–5761.

Boone, W. J. (2020). Rasch Basics for the Novice: Rasch Measurement. Springer, Singapore, pp. 9–30.

Brusilovsky P., Mill E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_1

Campbell, M. K.; Elbourne, D. R.; Altman, D. G.: CONSORT statement: extension to cluster randomised trials. *BMJ (Clinical research ed.)* 7441/328, pp. 702–708, 2004.

Dalipi, F.; Imran, A. S.; Kastrati, Z. (2018). MOOC Dropout Prediction Using Machine Learning Techniques: Review and Research Challenges. In *IEEE Global Engineering Education Conference*.

Deng, A.; Shi, X. (2016). Data-Driven Metric Development for Online Controlled Experiments. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM (ACM Digital Library), pp. 77–86.

Khine, M.S. Ed. (2020). Rasch Measurement. Springer Singapore, Singapore.

Kohavi, R.; Henne, R. M.; Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 959-967.

Kohavi, R.; Longbotham, R. (2017). Online Controlled Experiments and A/B Testing. In *Encyclopedia of machine learning and data mining*, 7(8), pp. 922-929.

Meier, C. (2019): KI-basierte, adaptive Lernumgebungen. In *Handbuch E-Learning*, pp. 1-21.

Paramythis, A.; Loidl-Reisinger, S. (2003). Adaptive learning environments and e-learning standards. In *2nd European Conference on e-Learning*.

Wijekumar, K. K.; Meyer, B. J. F.; Lei, P.: Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development* 6/60, pp. 987–1013, 2012.

Wong, B. T. M.; Li, K. C. (2018): Learning Analytics Intervention: A Review of Case Studies. In *2018 International Symposium on Educational Technology. ISET 2018*, pp. 178–182.

Xing, W.; Du, D. (2019). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. In *Journal of Educational Computing Research.* 57 (3), pp. 547–570. DOI: 10.1177/0735633118757015.