10-9-2023

Addressing Learners' Heterogeneity in Higher Education: An Explainable AI-based Feedback Artifact for Digital Learning Environments

Felix Haag
University of Bamberg, Germany, felix.haag@uni-bamberg.de

Sebastian A. Günther
University of Bamberg, Germany, sebastian.guenther@uni-bamberg.de

Konstantin Hopf
University of Bamberg, Germany, konstantin.hopf@uni-bamberg.de

Philipp Handschuh
Leibniz Institute for Educational Trajectories, Germany, philipp.handschuh@lifbi.de

Maria Klose
Leibniz Institute for Educational Trajectories, Germany, maria.klose@lifbi.de

See next page for additional authors

Authors

Felix Haag, Sebastian A. Günther, Konstantin Hopf, Philipp Handschuh, Maria Klose, and Thorsten Staake

Addressing Learners' Heterogeneity in Higher Education: An Explainable AI-based Feedback Artifact for Digital Learning Environments

Research Paper

Felix Haag[1], Sebastian A. Günther[1], Konstantin Hopf[1], Philipp Handschuh[2], Maria Klose[2], and Thorsten Staake[1]

[1] University of Bamberg, Information Systems and Energy Efficient Systems, Bamberg, Germany
{felix.haag, sebastian.guenther, konstantin.hopf, thorsten.staake}@uni-bamberg.de
[2] Leibniz Institute for Educational Trajectories, Bamberg, Germany
{philipp.handschuh, maria.klose}@lifbi.de

Abstract. Due to the advent of digital learning environments and the freedom they offer for learners, new challenges arise for students' self-regulated learning. To overcome these challenges, the provision of feedback has led to excellent results, such as less procrastination and improved academic performance. Yet, current feedback artifacts neglect learners' heterogeneity when it comes to prescriptive feedback that should meet personal characteristics and self-regulated learning skills. In this paper, we derive requirements from self-regulated learning theory for a feedback artifact that takes learners' heterogeneity into account. Based on these requirements, we design, instantiate, and evaluate an Explainable AI-based approach. The results demonstrate that our artifact is able to detect promising patterns in data on learners' behaviors and characteristics. Moreover, our evaluation suggests that learners perceive our feedback as valuable. Ultimately, our study informs Information Systems research in the design of future Explainable AI-based feedback artifacts that seek to address learners' heterogeneity.

Keywords: Digital Learning, Higher Education, Feedback, Explainable Artificial Intelligence, Counterfactual Explanations.

1 Introduction

Many institutions in higher education have responded to the growing popularity of Massive Open Online Courses (MOOCs) by either complementary or fully adapting face-to-face courses to digital learning environments. While the shift towards digital learning environments provides several advantages to instructors and learners (e.g., flexibility in scheduling teaching and learning sessions), this freedom also comes along with challenges for learners. More specifically, there is evidence that students with poor self-regulated skills suffer especially from this shift, potentially resulting in lower academic performance, whereas those with strong skills can cope well with a more digital setting (Kizilcec et al. 2017).

To support learners in digital learning environments, a plethora of artifacts came into existence that harnesses the potential of digital data traces recording their learning

18th International Conference on Wirtschaftsinformatik
September 2023, Paderborn, Germany

activities (Ifenthaler & Yau 2020). Recent examples in Information Systems (IS) research include learner support in the form of gamification elements (Benner et al. 2022, Leung et al. 2022), digital companions (Kruse et al. 2023), chatbots (Gupta & Chen 2022), and peer information (Li et al. 2021). Feedback (i.e., the provision of prescriptive information) thereby resembles an inherent way to support learners and represents a core mechanism for many of such IS artifacts. In general, feedback studies document intriguing behavioral effects on reducing procrastination or improving learning outcomes (see, e.g., Günther 2021, Theobald & Bellhäuser 2022).

Yet, due to their design, many of the feedback artifacts only support a specific type of learner (i.e., regarding self-regulated skills and prior knowledge of a topic). For instance, feedback in the mid of the semester designed to increase learning engagement might produce behavior change for students with low structuring skills (i.e., metacognition), while being less helpful for students with higher skills. This conjecture might also explain why studies often find mixed effects among different types of learners (see, e.g., Li et al. 2021, Leung et al. 2022). What is currently missing in IS research are approaches for feedback provision that directly address different types of learners regarding their heterogeneity and provide personalized instructions of a prescriptive nature.

We argue that technological advances in Machine Learning (ML) and Explainable Artificial Intelligence (XAI) can empower students by addressing learners' heterogeneity regarding their personal characteristics and Self-regulated Learning (SRL) skills through personalized instructions. ML has the capability to capture promising patterns in learning activities, learners' characteristics, and the interplay of both (i.e., activities and characteristics) that may resemble successful factors for individual academic success (Arnold & Pistilli 2012, Nguyen et al. 2020). Recent ML algorithms are able to model real-world relationships but usually hide the patterns detected in data from their users. XAI aims to overcome this issue by translating these patterns into a human-interpretable form (Meske et al. 2020). A potentially suitable XAI approach that allows to derive instructions of prescriptive nature from ML models is Counterfactual Explanations (CEs) (Fernández-Loría et al. 2022, Molnar 2019). CEs are motivated by the phenomenon that humans tend to judge events by the potential reasons why a desired event did not occur (Miller 2019), e.g., "Why did I get a poor grade in the exam instead of an excellent one?". To provide answers to such questions from the perspective of a ML model, CEs derive explanations in the form of required changes for a desired model outcome (Mothilal et al. 2020). Such explanations might also hold for real-world problems, e.g., "If I had spent more time in the learning platform, I wouldn't have had a poor grade in the exam". Hence, CEs have the potential to guide students' learning in digital environments through context-specific instructions that build on the learned patterns of the respective ML model. Therefore, this paper seeks to answer the following Research Question (RQ):

RQ: *How should an XAI-based artifact be developed that guides learning in digital environments through feedback provision while addressing learners' heterogeneity (e.g., regarding their characteristics and SRL skills)?*

To address this research question, we design, implement, and evaluate a user-centered feedback artifact that intends to leverage recent advances in ML and XAI to recognize learners' activities and characteristics and their interplay to derive meaningful feedback and therefore effective guidance during online learning. The overall goal of this study is to map the knowledge generated in SRL theory to an XAI-based artifact, which will contribute to the debate on providing effective learner support in digital environments (Gupta & Bostrom 2009). While our artifact opens a window to further research opportunities, we focus within this paper on the feasibility of such an approach.

Our paper is structured as follows: In the subsequent section, we outline the foundations of SRL theory to understand students' learning activities and introduce ML and XAI as concepts for learner guidance. Building on this background, we derive issues from phenomena documented in the literature that emerge for the design of such an artifact addressing learner heterogeneity. Relying on these issues, we synthesize requirements for the design of our artifact and use them to test our implementation. We finally highlight and discuss principles for future IS that implement XAI-based feedback to address subject heterogeneity in digital learning environments.

2 Background

2.1 Self-regulated Learning (SRL) Theory

Higher education poses many challenges to the learning processes of students. In contrast to secondary education, individuals have to rely more on their own self-regulated learning capabilities (i.e., the skill to plan, monitor and evaluate one's work and to control one's motivation and emotion) (Vosniadou 2020). Indeed, there is a stream of research under the notion of SRL that tries to understand which antecedents shape individuals' learning processes (see, e.g., Credé & Phillips 2011, Pintrich 2004, Wild & Schiefele 1994). Within the SRL framework, individuals differ with respect to cognitive, metacognitive, and resource management skills (Panadero 2017), which are fundamentally related to their context-specific choice of self-learning strategies. For instance, students with a higher level of time management skills tend to plan their learning better (e.g., by making and executing plans), while weaker skills could be related to self-regulation issues (e.g., procrastination). Therefore, to effectively guide students' SRL implies acknowledging these differences in students and also keeping students' motivation for learning high, as motivation is seen as a fundamental basis for SRL to take place (Pintrich 1999).

It is important to note that literature also provides a process-oriented view of how learners engage in SRL. More specifically, literature commonly distinguishes between several cyclical phases of SRL. For instance, Zimmerman & Moylan (2009) classify SRL activities into three consecutive phases: the forethought phase, the performance phase, and the self-reflection phase. In the forethought phase, learners plan tasks and set their learning goals, which also involves analyzing the value of the tasks at hand as well as assessing self-motivation beliefs (e.g., regarding their self-efficacy or outcome expectation in addressing a specific task) (Zimmerman & Moylan 2009). Subsequently, in the performance phase, learners execute different learning strategies (e.g., elaborate and repeat contents), as well as observe themselves to stay cognitively engaged and motivated (Panadero 2017). Finally, in the self-reflection phase, learners judge the progress towards their goals and make, based on the progress, attributions to their selves. For instance, in meeting or not meeting the learning goals of the forethought phase, learners might evaluate the suitability of the chosen learning strategies or reflect on themselves (e.g., eliciting positive or negative emotions associated with learning) (Zimmerman & Moylan 2009).

2.2 ML and XAI for Learner Guidance

Recent advances in the field of ML—one of the core technologies of Artificial Intelligence (AI)—have opened up new avenues for addressing the challenges of higher education learning in digital learning environments (Ouyang et al. 2022). Given that ML algorithms are capable of capturing complex patterns in large amounts of data (e.g., Berente et al. 2019), they could provide learners and instructors with additional insights. Examples include the early detection of students at-risk of failing an exam (Baneres et al. 2019) or the recommendation of courses (Bousbahi & Chorfi 2015).

Although the high predictive power of modern ML approaches is beneficial for application fields such as learning, models often come along with a lack of transparency and thus hide the patterns they detect in data from their users, leaving their potential to assist humans through the exposure of captured patterns untapped (Meske et al. 2020). The field of research that aims to improve the transparency of complex ML models is referred to as XAI (Barredo Arrieta et al. 2020).

A recent approach in XAI are CE methods that provide data-driven explanations on "what-if" questions (i.e., if a learner did a specific set of quizzes and videos, a model would estimate a better grade) (Fernández-Loría et al. 2022, Molnar 2019). CEs methods search for causal changes in features to shift a prediction to a desired output (Mothilal et al. 2020), thus allowing—on the premise that the model reflects real-world relationships (Štrumbelj et al. 2009)—for guidance involving the derivation of personalized instructions. Hence, CEs may be beneficial for higher education learning in three ways: First, by using CEs to demonstrate individually how to achieve a desired learning outcome (i.e., achieve more exam points). Second, as ML uses data on past learning behavior that has led to high and low learning success, the approach could suggest weaker students learning strategies from strong students. Exposing the behaviors of strong learners to potentially struggling learners has already been found to be effective in prior SRL studies (Davis et al. 2017). Third, by providing guidance on an individual level that is to some degree automated, it does not require much involvement of instructors.

3 Research Approach and Artifact Requirements

The research approach of our study is inspired by the principles of design science research (Hevner et al. 2004, Peffers et al. 2007). We thereby strive to tackle the challenge that current digital learning feedback artifacts do not account sufficiently for learners' heterogeneity (i.e., our problem specification). Following a nomothetic genre of design science inquiry (Baskerville et al. 2015), we identify issues from established knowledge (i.e., SRL and XAI literature) resulting from natural challenges in digital learning environments for the design of an XAI-based feedback artifact addressing learners' heterogeneity. Relying on these issues, we derive Meta-Requirements (MRs), implement the artifact, and evaluate it against measurable criteria. We finally outline design knowledge from artifact development and evaluation in the form of Design Principles (DPs) to address the general problem class at hand (i.e., effectively guide learners within the scope of digital learning environments). Given the limited scope of this paper, we briefly summarize the issues and related MRs below and outline the derived DPs inspired by Gregor et al. (2020) in Section 6.

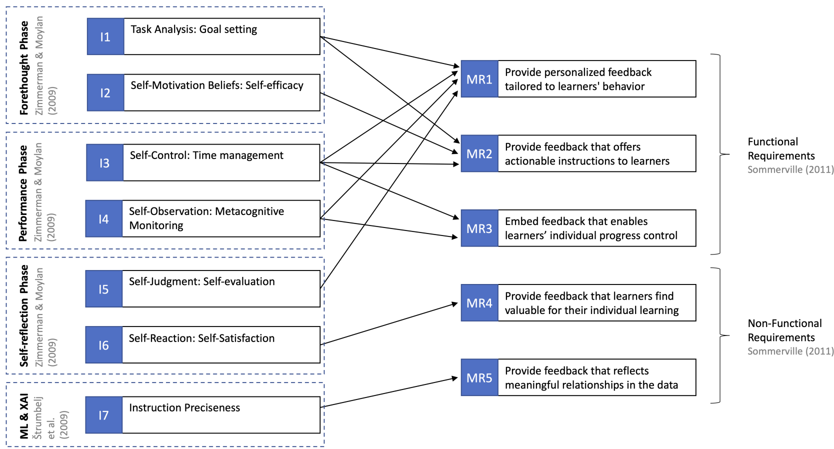

Figure 1. Issues (I1-I7) from SRL and XAI literature and MRs of the feedback artifact

We divide the MRs into functional (i.e., what the artifact must be able to do) and non-functional (i.e., properties of the artifact that need to be measurable) requirements according to Sommerville (2011), where we inherently consider the functional requirements for development and evaluate the non-functional ones against the artifact (Figure 1). In summary, we identified seven issues from literature. I1 to I6—derived from the cyclical phase model of self-regulatory feedback by Zimmerman & Moylan (2009)—aim to account for potential sources of heterogeneity in SRL. From these six issues, we derived four MRs. The first requirement (**MR1**), the provision of personalized feedback, should support students in specifying learning outcomes by informing them about the potential benefits of responding to feedback and allowing students to set specific task goals (I1) in accordance with their time management (I3). Moreover, the feedback should help them to metacognitively monitor their learning processes (I4), as well as to compare their progress to an evaluative standard (e.g., self-goals) (I5). Second, the artifact should meet the **MR2**, namely the provision of feedback in the form of actionable instructions to learners. Such instructions should stimulate goal setting (I1) and associated time management processes (I3), as well as lead to feelings of self-efficacy in case the tasks are perceived as manageable (I2). As the third requirement (**MR3**), the feedback of the artifact shall be embedded into a course overview page to enable learners' progress control. With the help of such progress control, learners who have deficits in time management (I3) and metacognitive monitoring (I4) can be supported, as the feedback might enable them to track their learning processes. Fourth, the artifact should provide feedback that learners find valuable for their individual learning (**MR4**), consequently be perceived as useful and be accepted by the learner (Strijbos et al. 2010). Valuable feedback is important for learners' self-satisfaction (I6), as learners prefer learning activities that previously led to satisfaction (compared to Bandura 2001). We evaluate this non-functional requirement within a user-based evaluation. The fifth requirement (**MR5**) stems from the XAI literature and refers to the precision of the instructions provided by the artifact (I7). More specifically, Štrumbelj et al. (2009) argue that the better the model recognizes domain concepts, the higher the quality of a data-driven explanation, and vice versa. The recognition of domain concepts, approximated through predictive performance, is therefore a prerequisite for the meaningful reflection of relationships in real-world data, and thus an essential non-functional system requirement for instruction quality.

4 Artifact Description

Our artifact is embedded into a digital learning environment and displays instructional feedback in the form of a component. The development builds on the aforementioned requirements and is designed to address learners' heterogeneity. Figure 2 illustrates an exemplary output of the component, displaying feedback to learners. The component consists of three elements and addresses MR1-MR3 (i.e., functional requirements): First, the introductory text, which refers to MR1 by pointing to the personalization of feedback and its tailoring to one's past online learning behavior. In doing so, the component explains that the additional implementation of instructions can lead to increased exam performance. Second, the three actionable instructions addressing MR2. We limit the instructions to three items because we argue that more instructions might lead to reactance in learners due to the considerable increase in additional workload. Third, check boxes learners can mark as "Done" to enable progress control, which addresses MR3.

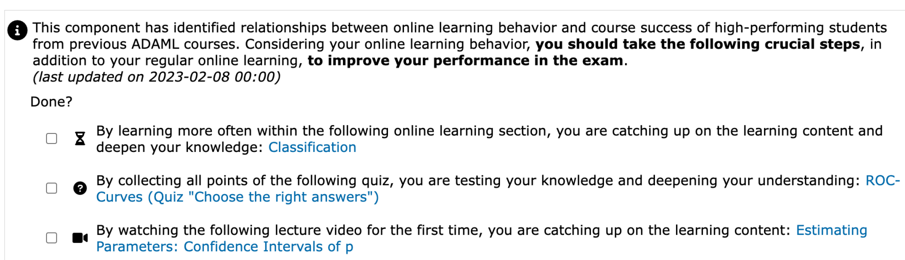

Figure 2. Exemplary component output for learners generated by the artifact

The quantitative modeling of our artifact to obtain feedback comprises four steps (Figure 3). These steps serve, on the one hand, for initially instantiating pre-trained ML models, and on the other, for generating feedback for an ongoing course. The paragraphs below describe each step in detail.

Data Acquisition. For data acquisition, our artifact collects behavioral data from an Open edX[1] online learning platform hosted by our university. On Open edX, students can access instructional videos, slides, and quizzes. The artifact considers extensive data on students' behavior and characteristics through built-in plugins in three ways. First, through the integration of the JavaScript library TimeMe.js[2] into the learning

[1] https://openedx.org/de/
[2] https://github.com/jasonzissman/TimeMe.js/

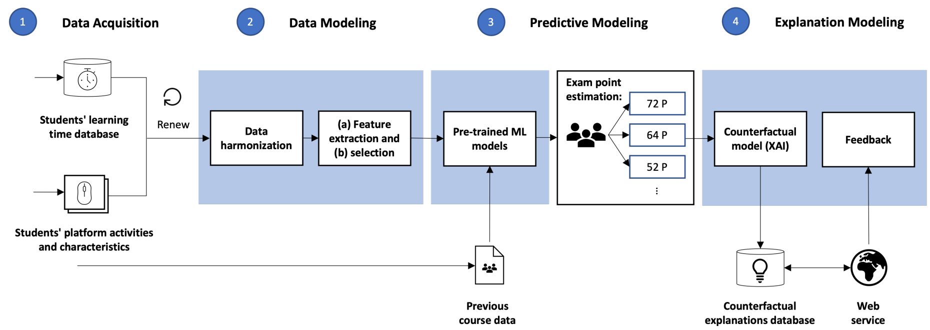

Figure 3. Quantitative modeling of the artifact

environment to more precisely record the *students' learning time* and save these data in an operational database. Second, by logging text files that record *students' platform activities and characteristics* in the form of user events on the edX environment (e.g., file downloads and videos watched etc.) and answers obtained from a built-in survey. Third, through the consideration of *previous course data* of prior semesters on students' learning time, platform activities, characteristics, and achieved exam points (i.e., the target variable) for a supervised training of the ML models.

Data Modeling. The functionality of our artifact initially performs *data harmonization* by merging the student learning time data and the log data containing the students' platform activities. Following the typical steps for model instantiation (Kühl et al. 2021), the resulting data set passes two successive preparation steps: *Feature extraction* (a) to obtain processable and meaningful features (i.e., predictor variables) that the artifact can feed into the ML algorithms. We list all features the artifact considers and mark features of type purely "predictive" with a "P" and all "actionable" features that we consider as part of our feedback artifact with "A" (Table 1). The rationale behind predictive-only features is that some may be misleading (e.g., video pauses) for feedback or hardly alterable but seem common for academic performance prediction (e.g., gender) due to their impact on students' learning (Hellas et al. 2018, Schoemer et al. 2021, Yukselturk & Bulut 2009). Finally, the artifact organizes the data set in an expanding window to ensure time-related predictive models and to consider only content and learner behavior that is relevant up to a certain point in time (e.g., for weekly granularity and feedback updates: Starting in the first semester week and gradually expanding the data set up to the final week). Moreover, we applied automatic *feature selection* (b) to capture the most important features and eliminate irrelevant ones. In doing so, our artifact includes a mutual information regression module (Kraskov et al. 2004, Ross 2014) that can cope with numerical and categorical features while maintaining interpretability.

Predictive Modeling. For the provision of feedback, we instantiate ML models to learn the relationship between learner behavior and characteristics, and course success. Our artifact predicts exam points achieved instead of the grade to avoid unnecessary dependent variable transformations and excludes exam retakers, and students with zero exam points from the target variable not to bias the ML models. We embed four well-recognized ML algorithms for predicting students' exam points that stem from different

Table 1. Variables used for prediction and actionable feedback

Group	*Variable*	*Description*	*Data type*[1]	*Type*[2]
Video	VideoViews.Total	Total number of video views	Int	A
	ViewsTotal.Complete	Total number of completely viewed videos	Int	A
	ViewsPerVideo{ID}	Indicates (unique) whether a specific video has already been watched	Int	A
	ViewsPerVideo{ID}.Complete	Indicates (unique) whether a video has already been watched completely	Int	A
	VideoPauses.Total	Total number of pauses while watching videos	Int	P
	VideoPlaybackSpeed.Total	Average speed with which the videos are played	Dbl	P
	VideoTranscriptHidden.Total	Total number of times a student hides the video transcript	Int	P
Quiz	QuizzesTaken.Total	Total number of quizzes taken	Int	A
	QuizTakenPerQuiz.{ID}	Indicates (unique) whether a specific quiz has already been taken	Int	A
	CorrectPerQuiz.{ID}	Total number of times a student achieved the maximum points of a specific quiz	Int	A
	CorrectRatioPerQuiz.{ID}	Ratio between the mean achieved points and the maximum points of a specific quiz	Dbl	P
File download	AssetDownload.{ID}	Indicates (unique) how often the slides of a specific tutorial/lecture have been downloaded	Int	A
Platform activity	Time.Total	Total time spent on the learning platform (incl. sum, SD, max)	Dbl	A
	HoursDuringDay.Total	Total time spent on the learning platform during the day (6:00 a.m. – 8:00 p.m.)	Dbl	P
	HoursDuringNight.Total	Total time spent on the platform during night	Dbl	P
	TimePerSection.{ID}	Total time spent per learning section	Dbl	A
	Requests.Total	Total number of HTTP request on the learning platform (incl. sum, SD, max)	Int/Dbl	P
	RequestsPerSection.{ID}	Total number of HTTP request per learning section	Int	P
	Sessions.Total	Total number of learning sessions (i.e., logins) on the learning platform (incl. sum, SD, max)	Int/Dbl	A
	SessionsPerSection.{ID}	Total number of learning sessions (i.e., logins) on the platform per learning section	Int	A
	BonusExerciseDone.{ID}	Indicates whether students have submitted a specific bonus assignment in the course to achieve extra points in the exam	Cat	P
Student characteristics	StudyProgram	Indicates the program (e.g., Information Systems) in which the student is enrolled	Cat	P
	Gender	Gender of the student (female, male, diverse)	Cat	P

[1] Int = Integer; Dbl = Double; Cat = Categorical
[2] A = Actionable; P = Predictive

categories of ML algorithms (e.g., Miguéis et al. 2018). We choose Extreme Gradient Boosting (XGBoost), CatBoost, Random Forest, and Support Vector Regression as estimators. The models are trained separately for each period within the expanding window. To reduce sampling bias, the artifact applies a stratified 80/20 train-test split according to students' exam points. Building on the training data set, we integrate the previously instantiated feature selection to build feature configurations using a percentage proportion p of the total features with $p \in \{20, 40, 60, 80, 100\}$ (i.e., $p\%$ of features with the highest score). Our artifact then trains models for each feature configuration and performs parameter tuning using a grid search applying a 5-fold cross-validation. The artifact automatically picks the model with the lowest prediction error on the test set according to the Root Mean Square Error (RMSE) to choose the best ML model for each period in time. Finally, our artifact fits *pre-trained models* on the whole data set (i.e., the entire previous course data, e.g., for a specific week) as the basis for generating CEs.

Explanation Modeling. To derive feedback for students, we rely on the XAI method Diverse Counterfactual Explanations (DiCE) introduced by Mothilal et al. (2020). The method seeks to propose diverse CEs (i.e., offer multiple explanations for behavioral change to increase exam points), which is desirable for real-world scenarios (Kusner et al. 2017). Given that our feedback is limited to three instructions, we employ the XAI method Shapley Additive Explanations (SHAP) (Lundberg & Lee 2017, Lundberg et al. 2020) for changes that affect more than three features. The *counterfactual model* stores the instructions for each student in an *CEs database*. Finally, the *feedback* component (Figure 2) displays instructions to learners by accessing the CEs database through a *web service*. We list the translation of CEs to actionable feedback in Appendix A.

5 Artifact Evaluation

Given the inherent fulfillment of functional requirements due to a corresponding development, we focus on the non-functional requirements for evaluation. Thereby, we follow the chronological order of the development process, starting with the evaluation of MR5 regarding the reflection of meaningful relationships followed by the evaluation of MR4. For artifact evaluation, we considered two distinct courses to obtain a more holistic picture across course levels: First, we employed feedback in a bachelor's IT Project Management course in the summer semester of 2022, and second, in a master's course on Data Analytics in the winter semester 2022/2023 from course week 7 on, respectively. As the artifact evolved between both iterations as part of the design cycle process, we had not yet included a feature selection method in the bachelor's course. Thus, we focus exemplarily on the master's course for MR5 and report the perceived value of feedback for both courses within the evaluation of MR4, as the wording of the feedback instructions remained consistent.

MR5 Evaluation. To assess the artifact's capability to reflect meaningful relationships, we considered data from three prior runs of the data analytics course summing up to $N = 118$ learners (Table 2). We combined the data from all three course runs and configured the artifact to account for weekly periods for training and evaluation of the ML models. Within the course, students can score 0–90 points in the exam.

Table 2. Overview of past data from three runs of the Data Analytics course

	Winter 2019/2020	*Winter 2020/2021*	*Winter 2021/2022*
Students (all course weeks)	46	39	33
Exam points mean (SD)	44.67 (24.40)	54.92 (20.60)	57.00 (17.43)
Learning time events	14753	23196	15309
Platform activity events	284334	196037	159080

Following Štrumbelj et al. (2009), we use predictive performance as a proxy metric and thereby leverage three well-established regression error measures: The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and RMSE (see Hastie et al. (2009) for details on the measures). Overall, we observe a decreasing error with increasing weekly data for all ML estimators on the test sets from week 9, which remains fairly constant for the following weeks (Figure 4). More specifically, we find the following error metric means (SDs) across all weeks and models the artifact relies up on: MAE $= 9.38\,(2.18)$, MAPE $= 23.98\,(6.09)$, and RMSE $= 11.90\,(2.64)$. To justify the usefulness of the ML approach, we additionally compare the residuals of the selected ML models with the naïve estimators (predicting with the respective Mean and Median of the train set), indicating the best performance for the mid (week 9) and end of the semester (week 17). To this end, we perform paired one-sided two-sample t-tests on the residuals of the test set (CatBoost vs. Median). The test results reveal a statistically significant decrease in the residuals of the ML models for the course week 9 ($t(21) = -2.55, p = 0.009, d = 0.54$) and 17 ($t(23) = -3.28, p = 0.002, d = 0.67$). Thus, we conclude that the ML approach is reasonable and assume that the artifact has the capability to model meaningful patterns of students' learning in the data.

MR4 Evaluation. We conducted online surveys in both courses (i.e., IT Project Management in summer 2022 and Data Analytics in winter 2022/2023) approximately

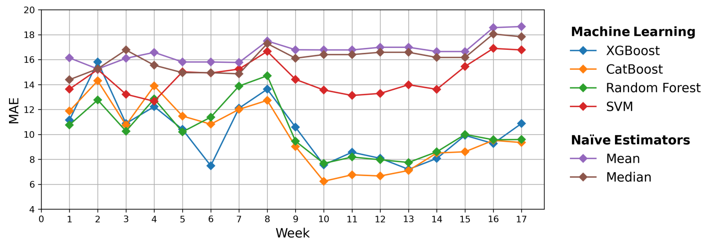

Figure 4. Predictive performance of the pre-trained ML models (MAE)

two weeks before the exam to assess the subjective value of our feedback artifact for students' learning after feedback had been exposed to a group of learners during the semester. To increase participation in the surveys, we incentivized the students with the opportunity to win an energy efficiency device worth approximately 60€. In summary, $N = 36$ learners responded to our surveys for which we used the sub-scales "usefulness" and "acceptance" of the scale "perceived adequacy of feedback" (Strijbos et al. 2010). Given that we have so far only exposed feedback to a subset of course participants, we collected responses on the sub-scales from 15 students, where 8 took the IT Project Management and 7 the Data Analytics course. Figure 5 displays the mean (SE) for both courses and each item. The results of our evaluation demonstrate that feedback recipients perceive the component similarly in both courses. On average, learners accept the component, perceive it as useful and helpful, and hardly reject or dispute it. However, we observe a high SE in the sub-scales due to the small sample size. We also find that 45.83% of the students that received feedback and participated in the final exam interacted with the component (i.e., through clicks on check boxes or instruction hyperlinks).

Based on our results, we assume that all artifact requirements (MR1-MR5) are fulfilled. Despite the encouraging results, the small sample size calls for further runs to refine the artifact as part of the design cycle process.

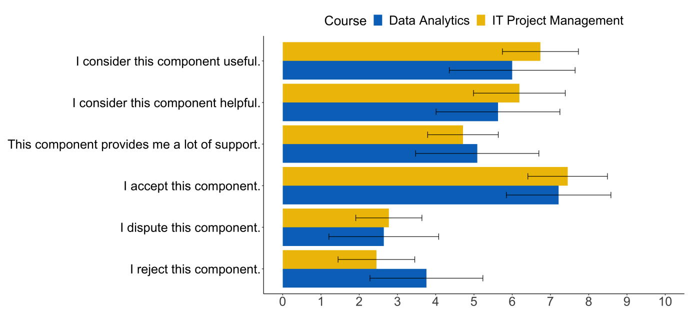

Figure 5. User-based evaluation results (items range from 1-fully disagree to 10-fully agree)

6 Discussion

To answer our RQ, we synthesize two DPs of the category "Design principles about an artifact" (Gregor et al. 2020) for the design of future XAI-based feedback artifacts to address learners' heterogeneity. First, we find that it is a key DP to *divide features associated with learners into predictive and actionable features to assure that learners can actually implement the displayed feedback instructions* (**DP1**). While the ML models consider every extracted feature to predict our dependent variable, the nature of our XAI-based feedback required us to curate the associated CEs outputs. Specifically, we divide the features into those that are predictive-only (i.e., hardly alterable characteristics such as study program or features that may be inappropriate) and those that are also actionable (i.e., from which the counterfactual method should obtain guiding instructions to stimulate, e.g., the repetition of specific videos). The purely predictive features thereby serve to represent learner heterogeneity properly by considering learners' activities that are potentially not suited for instructional feedback (e.g., "increase video playback speed" as the output from CEs) and rather immutable characteristics of learners. Second, from the artifact development and evaluation, we find that it is a key DP to *model learner-specific factors to assure that learners receive meaningful feedback instructions* (**DP2**). In contrast to a more technical application of ML for predictive analytics or XAI methods explaining why a model generates a particular output (e.g., for model inspection), XAI-based feedback that creates feedback addressing learners' heterogeneity requires embracing a socio-technical perspective. Thus, the design of such an artifact should not fully rely on correlations in the data, but also consider insights from validated theoretical models of human behavior. Our XAI-based feedback is inspired by SRL theory in the form that many of our ML features should stimulate and guide cognitive learning strategies (e.g., watch a specific video to repeat its content) and all phases of the SRL process (e.g., progress control in the performance phase). This may also hold for XAI-based feedback within other domains such as resource conservation (Wastensteiner et al. 2021), where theory from environmental psychology and human-computer interaction can guide the design of such artifacts.

7 Conclusion and Future Work

In this paper, we presented a novel feedback artifact that addresses learners' heterogeneity using XAI and CEs. The first evaluation highlights the feasibility of such an artifact and raises further questions on the effects of our feedback component. We anticipate that our feedback will have a positive impact on learners' online learning time, exam-taking behavior, and exam performance. In addition, we aim to investigate in future studies which type of learner benefits most from our feedback regarding academic success. In doing so, we plan to overcome the limitations of the first artifact iterations (e.g., small sample size). By shedding light on this form of feedback, our work will add to the understanding of how to effectively support students in digital learning environments, an area that has recently received much attention in the IS discipline.

8 Acknowledgements

This work was funded in part by "Stiftung Innovation in der Hochschullehre" within the project "Digitale Kulturen der Lehre entwickeln (DiKuLe)".

References

Arnold, K. E. & Pistilli, M. D. (2012), Course signals at Purdue: using learning analytics to increase student success, *in* 'Proceedings of the 2nd International Conference on Learning Analytics and Knowledge', LAK '12, Association for Computing Machinery, New York, NY, USA, pp. 267–270.

Bandura, A. (2001), 'Social cognitive theory: An agentic perspective', *Annual Review of Psychology* **52**(1), 1–26. tex.eprint: https://doi.org/10.1146/annurev.psych.52.1.1.

Baneres, D., Rodriguez, M. E. & Serra, M. (2019), 'An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course', *IEEE Transactions on Learning Technologies* **12**(2), 249–263.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020), 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion* **58**, 82–115.

Baskerville, R. L., Kaul, M. & Storey, V. C. (2015), 'Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production', *MIS Quarterly* **39**(3), 541–564. Publisher: Management Information Systems Research Center, University of Minnesota.

Benner, D., Schöbel, S. & Süess, C. (2022), Towards Gamified Conversational Agents for Self-Regulated Learning in Digital Education, *in* 'ICIS 2022 Proceedings', Copenhagen, Denmark.

Berente, N., Seidel, S. & Safadi, H. (2019), 'Research Commentary—Data-Driven Computationally Intensive Theory Development', *Information Systems Research* **30**(1), 50–64.

Bousbahi, F. & Chorfi, H. (2015), 'MOOC-Rec: A Case Based Recommender System for MOOCs', *Procedia - Social and Behavioral Sciences* **195**, 1813–1822.

Credé, M. & Phillips, L. A. (2011), 'A meta-analytic review of the motivated strategies for learning questionnaire', *Learning and Individual Differences* **21**(4), 337–346.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C. & Houben, G.-J. (2017), Follow the successful crowd: Raising mooc completion rates through social comparison at scale, *in* 'Proceedings of the 7th International Learning Analytics & Knowledge Conference', ACM Press, Vancouver, BC, Canada, pp. 454–463.

Fernández-Loría, C., Provost, F. & Han, X. (2022), 'Explaining data-driven decisions made by AI systems: The counterfactual approach', *MIS Quarterly* **46**(3), 1635–1660.

Gregor, S., Chandra Kruse, L. & Seidel, S. (2020), 'Research Perspectives: The Anatomy of a Design Principle', *Journal of the Association for Information Systems* **21**, 1622–1652.

Günther, S. A. (2021), The impact of social norms on students' online learning behavior: Insights from two randomized controlled trials, *in* 'LAK21: 11th International Conference on Learning Analytics & Knowledge', ACM, Irvine, CA, USA, pp. 12–21.

Gupta, S. & Bostrom, R. (2009), 'Technology-Mediated Learning: A Comprehensive Theoretical Model', *Journal of the Association for Information Systems* **10**, 686–714.

Gupta, S. & Chen, Y. (2022), 'Supporting inclusive learning using chatbots? A chatbot-led interview study', *Journal of Information Systems Education* **33**(1), 98–108.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York.

Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C. & Liao, S. N. (2018), Predicting academic performance: a systematic literature review, *in* 'Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education', Larnaca, Cyprus, pp. 175–199.

Hevner, March, Park & Ram (2004), 'Design Science in Information Systems Research', *MIS Quarterly* **28**(1), 75.

Ifenthaler, D. & Yau, J. Y.-K. (2020), 'Utilising learning analytics to support study success in higher education: a systematic review', *Educational Technology Research and Development* **68**(4), 1961–1990.

Kizilcec, R. F., Pérez-Sanagustín, M. & Maldonado, J. J. (2017), 'Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses', *Computers & Education* **104**, 18–33.

Kraskov, A., Stögbauer, H. & Grassberger, P. (2004), 'Estimating mutual information', *Physical Review E* **69**(6), 066138.

Kruse, L. C., Bergener, K., Conboy, K., Lundström, J. E., Maedche, A., Sarker, S., Seeber, I., Stein, A. & Tømte, C. E. (2023), 'Understanding the digital companions of our future generation', *Communications of the Association for Information Systems* **52**(1), 22.

Kusner, M. J., Loftus, J., Russell, C. & Silva, R. (2017), Counterfactual fairness, *in* 'Advances in Neural Information Processing Systems', Vol. 30, Curran Associates, Inc.

Kühl, N., Hirt, R., Baier, L., Schmitz, B. & Satzger, G. (2021), 'How to Conduct Rigorous Supervised Machine Learning in Information Systems Research: The Supervised Machine Learning Report Card', *Communications of the Association for Information Systems* **48**(1), 589–615.

Leung, A. C. M., Santhanam, R., Kwok, R. C.-W. & Yue, W. T. (2022), 'Could gamification designs enhance online learning through personalization? lessons from a field experiment', *Information Systems Research* .

Li, Z., Wang, G. & Wang, H. (2021), 'Peer effects in competitive environments: Field experiments on information provision and interventions', *MIS Quarterly* **45**(1), 163–191.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. & Lee, S.-I. (2020), 'From local explanations to global understanding with explainable AI for trees', *Nature Machine Intelligence* **2**(1), 56–67.

Lundberg, S. M. & Lee, S.-I. (2017), A unified approach to interpreting model predictions, *in* 'Advances in neural information processing systems', Vol. 30, Curran Associates, Inc.

Meske, C., Bunde, E., Schneider, J. & Gersch, M. (2020), 'Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities', *Information Systems Management* pp. 1–11.

Miguéis, V., Freitas, A., Garcia, P. J. & Silva, A. (2018), 'Early segmentation of students according to their academic performance: A predictive modelling approach', *Decision Support Systems* **115**, 36–51.

Miller, T. (2019), 'Explanation in artificial intelligence: Insights from the social sciences', *Artificial Intelligence* **267**, 1–38.

Molnar, C. (2019), *Interpretable machine learning - A Guide for Making Black Box Models Explainable*.

Mothilal, R. K., Sharma, A. & Tan, C. (2020), Explaining machine learning classifiers through diverse counterfactual explanations, *in* 'Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency', ACM, Barcelona Spain, pp. 607–617.

Nguyen, A., Gardner, L. & Sheridan, D. (2020), 'Data Analytics in Higher Education: An Integrated View', *Journal of Information Systems Education* **31**(1), 61–71.

Ouyang, F., Zheng, L. & Jiao, P. (2022), 'Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020', *Education and Information Technologies* **27**(6), 7893–7925.

Panadero, E. (2017), 'A review of self-regulated learning: Six models and four directions for research', *Frontiers in Psychology* **8**.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007), 'A Design Science Research Methodology for Information Systems Research', *Journal of Management Information Systems* **24**(3), 45–77.

Pintrich, P. R. (1999), 'The role of motivation in promoting and sustaining self-regulated learning', *International Journal of Educational Research* **31**(6), 459–470.

Pintrich, P. R. (2004), 'A conceptual framework for assessing motivation and self-regulated learning in college students', *Educational Psychology Review* **16**(4), 385–407.

Ross, B. C. (2014), 'Mutual Information between Discrete and Continuous Data Sets', *PLoS ONE* **9**(2), e87357.

Schoemer, D., Laumer, S., Wilbers, K., Wolbring, T., Weigert, J. & Treischl, E. (2021), Data-driven Student Advisory and Potential Direct Discrimination: A Literature Review on Machine Learning for Predicting Students' Academic Success, *in* 'AMCIS 2021 Proceedings', Montreal, Canada.

Sommerville, I. (2011), *Software engineering*, 9th ed edn, Pearson, Boston. OCLC: ocn462909026.

Strijbos, J.-W., Narciss, S. & Dünnebier, K. (2010), 'Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?', *Learning and instruction* **20**(4), 291–303. Publisher: Elsevier.

Theobald, M. & Bellhäuser, H. (2022), 'How am i going and where to next? elaborated online feedback improves university students' self-regulated learning and performance', *The Internet and Higher Education* **55**, 100872.

Vosniadou, S. (2020), 'Bridging secondary and higher education. the importance of self-regulated learning', *European Review* **28**(S1), S94–S103.

Wastensteiner, J., Weiss, T., Haag, F. & Hopf, K. (2021), Explainable AI for Tailored Electricity Consumption Feedback – an Experimental Evaluation of Visualizations, *in* 'ECIS 2021 Research Papers', 29, Marrakech, Morocco.

Wild, K.-P. & Schiefele, U. (1994), 'Lernstrategien im studium: Ergebnisse zur faktorenstruktur und reliabilität eines neuen fragebogens', *Zeitschrift für Differentielle und Diagnostische Psychologie* **15**(4), 185–200.

Yukselturk, E. & Bulut, S. (2009), 'Gender differences in self-regulated online learning environment', *Journal of Educational Technology & Society* **12**(3), 12–22. Publisher: JSTOR.

Zimmerman, B. J. & Moylan, A. R. (2009), Self-regulation: Where metacognition and motivation intersect, *in* 'Handbook of metacognition in education', Routledge, pp. 311–328.

Štrumbelj, E., Kononenko, I. & Šikonja, M. R. (2009), 'Explaining instance classifications with interactions of subsets of feature values', *Data & Knowledge Engineering* **68**(10), 886–904. Publisher: Elsevier.

Appendix A Feedback Instructions

Table 3. Translation of actionable variables to feedback instructions

Actionable variable	*Feedback (initial instruction)*	*Feedback (repeated instruction)*
VideoViews.Total	By watching lecture videos for the first time, you are catching up on the learning content.	By watching more lecture videos, you are catching up on the learning content and deepen your understanding.
ViewsTotal.Complete	By watching lecture videos for the first time completely, you are catching up on the learning content.	By watching more lecture videos completely, you are catching up on the learning content and deepen your understanding.
ViewsPerVideo{ID}	By watching the following lecture video for the first time, you are catching up on the learning content: {Video Name}	You already know the learning content of the following lecture video. Watch this lecture video again to deepen your understanding: {Video Name}
ViewsPerVideo{ID} .Complete	You are still missing parts of a lecture video. Watch the following lecture video for catching up on the missing parts: {Video Name}	You already know the learning content of the following lecture video. Watch this lecture video again completely to deepen your understanding: {Video Name}
QuizzesTaken.Total	By taking quizzes for the first time, you are testing your knowledge on the learning content.	By taking more quizzes, you are testing and monitoring your knowledge on the learning content.
QuizTakenPerQuiz.{ID}	By taking the following quiz for the first time, you are testing your knowledge: {Quiz Name}	You have already tested your knowledge with the following quiz. Try the quiz again for monitoring your knowledge: {Quiz Name}
CorrectPerQuiz.{ID}	By collecting all points of the following quiz, you are testing your knowledge and deepening your understanding: {Quiz Name}	You have already tested your knowledge with the following quiz. Try to solve the quiz again correctly for monitoring your knowledge: {Quiz Name}
AssetDownload.{ID}	By downloading and working through the following slides, you are catching up on the learning content: {File Name}	You have already downloaded these slides. Open and work through them again to deepen your understanding: {File Name}
Time.Total	You have already spent valuable learning time on the edX platform. By further increasing your online learning time, you are catching up on the learning content and deepen your knowledge.	
TimePerSection.{ID}	You have already spent valuable learning time on the edX platform. By spending more online learning time in the following learning section, you are catching up on the learning content and deepen your knowledge: {Section Name}	
Sessions.Total	By learning more often within the edX platform, you are catching up on the learning content and deepen your knowledge.	
SessionsPerSection.{ID}	By learning more often within the following online learning section, you are catching up on the learning content and deepen your knowledge: {Section Name}	