Klose, Maria; Handschuh, Philipp; Steger, Diana; Artelt, Cordula

Embracing the challenge : Predicting self-testing in non-formal online courses using machine learning

Date of secondary publication: 12.12.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-112132x

Primary publication

Klose, Maria; Handschuh, Philipp; Steger, Diana; u. a. (2026): Embracing the challenge : Predicting self-testing in non-formal online courses using machine learning, in: Computers & education : an international journal, Amsterdam [u.a.]: Elsevier Science, Vol. 242, Nr. 105507, pp. 1–13, doi: 10.1016/j.compedu.2025.105507.

Computers & Education 242 (2026) 105507

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Embracing the challenge: Predicting self-testing in non-formal online courses using machine learning

Maria Klose [a,*], Philipp Handschuh [a], Diana Steger [a], Cordula Artelt [a,b]

[a] Leibniz Institute for Educational Trajectories (LIfBi), Bamberg, Germany
[b] Department of Longitudinal Educational Research, University of Bamberg, Germany

ARTICLE INFO

Keywords:
Online learning
Technology-enhanced learning
Lifelong learning
Self-testing
Machine learning

ABSTRACT

Technology-driven advancements have made adaptive and interactive learning techniques more accessible. Online courses increasingly integrate self-tests that offer automated and immediate feedback. Self-tests can help learners to identify knowledge gaps and to reinforce their retention and comprehension. However, not all learners readily use self-tests, raising the question of which factors may impact learners' engagement with self-tests. The present study focused on non-formal education, covering 45 online courses offered by Bavarian universities. Analyses were based on a sample of $N = 1261$ participants aged 16–84 years. We used a machine learning approach to predict learners' engagement with self-tests and to identify important influencing factors. Therefore, we included 50 predictor variables in an elastic net regression to explore the role of learner- and course-related characteristics. The predictor variables were drawn from self-report, process, and meta data. Overall, learners differed substantially in their self-testing behavior. The prediction model explained 11 % of the variance in learners' engagement with self-tests. Despite the model's modest explanatory power, the analysis identified potentially relevant predictors. The two most important predictor variables were learner commitment and the intention to obtain a confirmation of participation. Accordingly, course designers might implement extrinsic incentives—such as confirmations of participation—as a potentially useful strategy to encourage learners' engagement with self-tests. From a methodological perspective, the study highlights the importance of using appropriate statistical methods—such as machine learning algorithms—to understand complex learning behaviors.

1. Introduction

In the age of digitalization, education is undergoing a rapid transformation, offering new opportunities for lifelong learning (Mozelius et al., 2024). Technology-enhanced learning facilitates access to formal, informal, and non-formal education across the lifespan (Jaldemark et al., 2021). Among these, non-formal online courses enable learners to easily access organized learning materials, but without formal accreditation (Schwier & Seaton, 2013). These courses typically lack external control, for example by instructors or graded performance-based systems. Thus, greater demands are placed on learners' self-regulation skills (Kizilcec et al., 2017), which include goal setting, strategy use, progress monitoring, and time management (Zimmerman, 2002). In general, one effective learning technique to autonomously monitor and regulate the learning process is the engagement with self-tests (i.e.,

* Corresponding author.
E-mail address: maria.klose@lifbi.de (M. Klose).

https://doi.org/10.1016/j.compedu.2025.105507
Received 28 February 2025; Received in revised form 13 October 2025; Accepted 15 November 2025
Available online 20 November 2025

completing tests as a no-stakes practice and on one's own; Dunlosky et al., 2013). However, although it is known that self-tests are powerful tools to improve learning performance in both formal and non-formal learning contexts (Rowland, 2014), not all learners readily engage in self-testing, as it might be seen as a tedious necessity (Dunlosky et al., 2013). With the present study, we aim to better understand which learners persist in engaging with self-tests within non-formal online courses, which can be regarded as prototypical self-regulated learning environments.

1.1. *Self-tests as formative assessments in non-formal online courses*

Self-tests are a form of formative assessment, which accompany the learning process and support learning by providing immediate feedback on the correct solutions (e.g., Dolin et al., 2018; Wiliam & Black, 1996). Unfortunately, self-testing can be perceived negatively, as many learners mistakenly equate self-tests with high-stakes summative assessments, which are often seen as tools for judgment rather than learning (Dunlosky et al., 2013), as summative assessments are used for the purpose of grading or certification, and are given mid-term or at the end of a course (e.g., Bloom et al., 1971; Dixson & Worrell, 2016). Consequently, learners shy away from self-testing as a learning technique and tend to use less effective learning techniques (e.g., passive review methods like rereading and highlighting) that require less time and effort, because they misinterpret the mental effort of self-testing as a sign of poor learning (Kirk-Johnson et al., 2019).

Learners' perceptions of self-testing as an inefficient technique stand in contrast to extensive evidence demonstrating its robust effectiveness in promoting long-term retention across numerous studies in formal learning contexts (e.g., Adesope et al., 2017; Yang et al., 2021). Importantly, comparable effects of self-testing have also been observed in non-formal learning environments, indicating that its effectiveness is not confined to institutional settings (Rowland, 2014). Tests do not only assess knowledge to identify which course content has or has not been understood, but they also directly enhance learning by initiating retrieval processes when learners are taking tests (i.e., *testing effect*; Karpicke, 2017). Therefore, the benefits of testing for learning are manifold (Roediger et al., 2011). For example, testing identifies gaps in knowledge, improves transfer of knowledge to new contexts, and triggers active retrieval which fosters later retention.

Despite these well-documented benefits, self-tests may also entail unintended negative side effects, such as false memory effects. Especially in the case of recognition tests (e.g., multiple-choice tests, true/false statements, matching exercises), the correct answer is presented along with distractors that provide wrong statements or erroneous information. This might ultimately lead to the acquisition of false knowledge (Marsh et al., 2007) if learners memorize the wrong answer. This so-called *negative suggestion effect* (Roediger & Karpicke, 2006) can be diminished by providing corrective feedback (Kang et al., 2007). It might be even eliminated if learners receive the feedback promptly after taking the test (Roediger & Karpicke, 2006), provided that they learn from the feedback and understand the principles underlying their errors. Therefore, it is not only important for learners to receive immediate feedback, but also to interact with the feedback to correct their errors to overcome potential negative effects of testing (i.e., cognitive engagement).

In general, learning engagement is commonly conceptualized across three interrelated dimensions: affective, behavioral, and cognitive engagement (Bond et al., 2020). *Affective engagement* includes aspects like interest and a positive attitude towards learning, *behavioral engagement* encompasses participation and involvement in learning activities, and *cognitive engagement* involves self-regulation and investment in learning (Appleton et al., 2008; Fredricks et al., 2004). Our learning behavior of interest, engagement with self-tests, goes beyond learners participating in the learning process by taking the self-tests (i.e., behavioral engagement; Fredricks et al., 2004). Learners must not only participate in self-tests, but rather invest sustained effort in using self-testing as a self-regulated learning technique over time to achieve mastery of the learning material (Lawson & Lawson, 2013).

1.2. *Who engages with self-tests and under what conditions?*

A substantial body of research has already examined whether the effectiveness of testing as a learning technique is moderated by interindividual differences, concluding that testing is beneficial for all learners irrespective of, for example, personality traits or working memory capacity (Bertilsson et al., 2021), cognitive ability (Jonsson et al., 2021), age (Pastötter & Bäuml, 2019), and motivational or emotional dispositions (Glaser & Richter, 2023). Apart from learner characteristics, testing has proven effective across different learning environments (Rivers, 2021)—including content areas, education levels (Agarwal et al., 2021), and study settings (Adesope et al., 2017). Together, these findings underscore the robustness of self-testing as a learning technique that benefits a wide range of learners across diverse contexts.

However, to benefit from the testing effect, learners must first engage in self-testing. Accordingly, our focus is not on evaluating whether engaging with self-tests improves course performance, but rather on examining whether learners utilize self-tests as a learning technique at all, including the persistence in interacting with corrective feedback until tests are answered correctly as a proxy for cognitive engagement. In the context of the present study, the key question is how learners can be encouraged to adopt self-testing as a learning technique by identifying relevant predictors for the engagement with self-tests. Given the ongoing debate about which learners engage in self-testing, a comprehensive approach that considers both learner and course characteristics, along with multiple data sources, is essential (Carpenter, 2023).

In the process of self-regulated learning, learners are free to choose to use self-testing as a learning technique (Pintrich, 2000). Prior research has shown that this choice is associated with various learner characteristics, including socio-demographic factors, cognitive, and non-cognitive aspects (e.g., Bertilsson et al., 2024; Fellman et al., 2020). In particular, motivation plays a central role: learners who are more motivated to engage in learning activities are also more likely to use self-tests (Bertilsson et al., 2024), emphasizing the importance of context-specific motivational factors for the engagement with self-tests. Beyond motivation, learners' interest in the learning task (Zepeda et al., 2020) and their ability to employ self-regulation strategies to maintain effort regulation towards self-testing (De Bruin et al., 2023) are further prerequisites for continued engagement. Moreover, learning behavior and engagement levels are influenced by the initial reasons for enrolling in a voluntary (online) course (Kizilcec & Schneider, 2015), which can also impact the adoption of self-testing as a learning technique. Overall, these findings suggest that a variety of learner characteristics are associated with engagement with self-tests, which is why the joint investigation of all of the aforementioned characteristics may adequately depict the complex interrelations.

In addition to learner characteristics, contextual aspects of the learning environment that influence learners' engagement with self-tests must also be considered (Zepeda et al., 2020). Research on the testing effect has primarily been conducted in formal learning contexts (i.e., in the classroom or in standardized laboratory experiments; e.g. Adesope et al., 2017; Yang et al., 2021). Consequently, there is limited research on the effects of different learning contexts. Understanding the impact of contextual variables like course characteristics becomes even more crucial when investigating authentic, real-life learning scenarios, especially since non-formal online courses vary greatly in their course design (Maya-Jariego et al., 2020). For instance, course duration can influence perceived workload and, in turn, learners' sustained engagement (Eriksson et al., 2017; Goopio & Cheung, 2021), especially since such courses are typically taken in participants' leisure time.

1.3. *The present study*

Not all learners take advantage of self-tests as a learning opportunity (Carpenter et al., 2022). Especially in a non-formal online learning context, learners have to decide for themselves whether to invest effort in the voluntary option of self-testing. Thus, it is crucial to understand which learners are willing to commit to self-testing and under which conditions they persist until they can answer the tests correctly. Accordingly, in the present study, we aim to predict the learners' engagement with self-tests in non-formal online courses.

However, one's willingness to engage with self-tests potentially hinges on a variety of different factors. Drawing on the conclusion from a recent review on effective learning with retrieval practice (Carpenter et al., 2022), we investigated learners' engagement with self-tests a) within a real (online) learning scenario, b) based on a comprehensive dataset that simultaneously covers learner- and course-related characteristics, and c) using multiple data sources. Thereby we want to explore potential factors that may impact learners' engagement with self-tests, providing potential strategies to make self-testing more accessible to a broader range of learners. As we draw on a large number of predictor variables, we use a machine learning approach to appropriately model the extent to which learners engage with self-tests as a learning technique throughout an online course.

2. Methods

Following common open science practices, all supplemental study materials (i.e., codebook, syntax, data, and supplemental figures and tables) are available online within the Open Science Framework (Center for Open Science, 2025): https://osf.io/5za3d/.

2.1. *Design and participants*

The present study was part of a larger research project with a longitudinal design. For the current analyses, we rely on cross-sectional questionnaire data only, which was collected at the first assessment (for more detailed information see Klose et al., 2024). These data were collected between October 2021 and April 2022 via an online survey implemented within 45 non-formal online courses offered by German, particularly Bavarian, universities (OPEN vhb; Virtuelle Hochschule Bayern, 2025). Data collection was conducted in accordance with the ethical guidelines established by the German Society for Online Research, the declaration of Helsinki, and the General Data Protection Regulation (GDPR). After course enrollment, a link to the survey was provided on the first page, allowing learners to participate voluntarily. Participants actively approved their involvement by agreeing to take part, and their anonymity was fully guaranteed. Ethical approval was not required under local legislation. Participants answered questions on socio-demographic information, enrollment intentions (*OLEI* Scale; Kizilcec & Schneider, 2015), situational motivation (*SIMS*; Guay et al., 2000), situational interest (*SIS*; Linnenbrink-Garcia et al., 2010), and self-regulation (*LIST-K*; Klingsieck, 2018). Table S1 provides a comprehensive overview of the survey content compilation, detailing the specific subscales utilized, any adaptations made, the number of items, and the response format. The online courses were available to anyone interested at no cost, following a self-paced format (i.e., enrollment at any time, with all course materials available from the outset, and no set deadlines). Moreover, the courses were taught in German and covered a broad range of subjects, for example social sciences and humanities, science, technology, engineering, and mathematics (STEM), health science, and economics. Each course included self-testing opportunities that were required for learners to receive a confirmation of participation, which could be downloaded at the end of the course. However, this document is no official certification or degree but simply verifies attendance. The required percentage of correct test answers per course ranged from 50 % to 100 % ($M = 77$ %; $SD = 13$ %). In addition, the courses' length varied with respect to the estimated time needed to complete the course (range 9–60 h; $M = 18.88$; $SD = 9.38$), and its number of chapters (range 3–16; $M = 7.49$; $SD = 2.75$) and lessons (range 11–143; $M = 56.70$; $SD = 27.44$).

The online survey resulted in a total of 1307 cases, of which 1126 (96.5 %) were included in the final analyses and consisted of 1205 unique subjects (95.6 %; learners were allowed to participate in the survey for each course they attended, so that multiple participation in the study was possible if they were enrolled in several courses). No sample-size-dependent stopping rule for data collection was specified beforehand, as the planned analyses require comparatively large sample sizes for stable estimations. Using machine learning algorithms, issues of underpower are generally more discussed than issues of overpower, since limited sample sizes are for example problematic for pattern recognition (Vabalas et al., 2019). Generally, our current sample size falls within the range of sample sizes reported by studies using comparable methods (e.g., Greenwood et al., 2020; Wang et al., 2024). Table S2 displays the detailed data cleaning procedure and the participant dropout at different stages due to missing data or validation checks. In the final sample, 68.0 % of the participants reported being female (30.9 % male, and 1.1 % non-binary or non-response). The average age was 37.0 years ($SD =$ 14.30, range 16–84 years), encompassing all age groups from older adolescents to adults of retirement age (>65 years; see Fig. S1 for the age distribution). The sample predominantly consists of individuals with academic backgrounds: 47.8 % had an university degree, 34.1 % had a secondary school degree qualifying for university (i.e., academic-track: *Gymnasium* or *Fach-/Berufsoberschule*), 10.4 % had a middle secondary school degree (i.e., intermediate-track: *Realschule*), 2.6 % had a lower secondary school degree (i.e., vocational-track: *Hauptschule*), and 5.1 % had a different educational background (i.e., no degree, other school degree, still in school, or non-response). In terms of the current employment status of the sample, the distribution was as follows: 58.1 % were employed, 23.6 % were school/university students, 4.3 % were apprentices, 5.0 % were retired and 9.0 % had a different employment status (i.e., seeking employment, unemployed, or non-response). All participants were fluent in German. While our sample is not representative of a broader population, its socio-demographic composition is consistent with those observed in other studies on non-formal education in Germany. In particular, participation often skews towards younger adults and individuals with higher levels of education (Kruppe & Baumann, 2019). The presence of older participants (ages 50–64 and above) also reflects broader trends in lifelong learning participation (see Bilger & Strauβ, 2022 for continuing education behavior in Germany). Table S3 provides descriptive information on courses and participants separately by course to provide an overview over the hierarchical data structure (i.e., learners nested in courses).

2.2. Measures

The present study followed an exploratory, data-driven, machine learning approach (Yarkoni & Westfall, 2017) and therefore included a broad set of predictor variables across different data types to maximize the opportunity for detecting relevant patterns (Breiman, 2001). To predict learners' engagement with self-tests, three types of data were considered: *meta data* that was coded from openly available course information as course-related characteristics, as well as *process data* from the online survey, and *questionnaire data* (i.e., self-report scales from the online survey) as learner-related characteristics. While not all predictors were derived from theory, their inclusion was justified by their plausible relationship to the dependent variable. Concerns such as multicollinearity or overfitting were addressed by the choice of appropriate statistical modeling techniques. Table S1 summarizes all 50 variables used as predictors with a detailed description of each variable including descriptive statistics. For categorical variables, the category of interest is specified for dummy-coding. For each subscale, Cronbach's alpha is indicated as a measure of internal consistency.

We operationalized learners' engagement with self-tests using the percentage of correct test answers, derived from log data transmitted by the *Virtuelle Hochschule Bayern* (*vhb)*, the provider of the non-formal online courses. This percentage reflects the correct test answers recorded throughout the course, regardless of the frequency of learners' test attempts (i.e., the proportion of self-tests that learners answered correctly at least once). Consequently, learners with no correct test answers included those who did not take the self-test option at all, as well as those who took the tests but answered incorrectly. Since these tests were designed as formative assessments—characterized by no-stakes testing situations, immediate feedback on performance, and unlimited test attempts—we argue that the percentage of correct test answers serves as a proxy for the learners' cognitive engagement (i.e., investing effort until a part of mastery of learning content is achieved). It does not, however, reflect actual course performance, as learners' proficiency goes beyond the percentage of correct test answers (Abbakumov et al., 2020). Each course offered at least three self-tests ($M = 26$; $SD = 32$; max. $=$ 179), that were distributed throughout the course to accompany the learning process. The tests covered a variety of formats, including multiple-choice quizzes, matching tasks, true/false statements, and fill-in-the-blank exercises, all designed to capture course material without requiring prior knowledge. While participation in the self-tests was voluntary, completion of a certain portion of the self-tests was a requirement for confirmation of participation.

2.3. Statistical analyses

2.3.1. Score computation and missing data

We aggregated scales using mean scores for each subscale of the questionnaires. Where necessary, we recoded reversed items before aggregation. Overall, the amount of missing data for metric predictors was small, with a maximum missingness of 7.5 % per variable and an average missingness of 2.9 % across all variables. Given this low level of missingness, we imputed missing values for metric predictors using the *k*-nearest neighbor algorithm, which is considered a robust and sensitive method for missing value estimation for data sets with less than 10 % missing entries (Troyanskaya et al., 2001). All metric predictor variables were standardized before entered in the machine learning models.

For categorical variables, we applied dummy coding to include them in the analyses as dichotomous predictors using one category of interest as specified in Table S1. Missing data was coded as 0 and therefore grouped together with all other categories and treated as a reference category. The amount of missing data for dichotomous predictors was slightly higher with an average missingness of 5.3 %. The variables with the largest proportion of missingness were enrollment intentions with a maximum missingness of 25.1 % for the statement "course offered by prestigious professor". Out of 14 enrollment intentions, 12 fall below the rule-of-thumb limit of maximum 10 % missingness. We decided to include all dichotomous predictors in the final analyses because we treated missing data the same as disagreement (i.e., treated as reference category) since almost all missing data were due to the deliberate decision to choose the "no answer" option rather than dropping out of the survey.

2.3.2. *Elastic net regression*

To predict the engagement with self-tests, we used elastic net regression—a statistical method that builds a prediction model while automatically reducing the influence of less important variables. Elastic net regression is a shrinkage and variable selection method (Zou & Hastie, 2005) which combines ridge regression (Hoerl & Kennard, 2000) and LASSO regression (i.e., least absolute shrinkage and selection operator; Tibshirani, 1996) as forms of regularization to avoid overfitting. Since ridge regression cannot produce a parsimonious model, and lasso regression cannot handle correlated predictor variables, elastic net regression creates a useful compromise via the regularization parameters λ and α (Zou & Hastie, 2005): λ controls the overall strength of penalty applied to the model, and α determines the mix between ridge and lasso penalties (see Helwig, 2017 for more details). This combination allows elastic net regression to handle correlated predictors by shrinking coefficients of correlated predictors towards each other, and less important coefficients towards 0 (ridge), while still achieving a sparse prediction model solution through variable selection (lasso) at the same time (Friedman et al., 2010).

To ensure unbiased model evaluation, we split the full dataset into 80 % of the courses for training and 20 % of the courses for independent testing of the prediction model. Given the hierarchical structure of our data (i.e., learners nested in courses), we employed a blocked cross-validation approach (Roberts et al., 2017) to prevent bias in performance estimates (Pargent et al., 2023). Consequently, all observations belonging to one online course were assigned either to the training set or the test set (i.e., blocked resampling). The consideration of the aforementioned multilevel structure in our analyses was supported by an intraclass correlation coefficient (ICC) for a null multilevel model, as the estimated ICC = .13 indicated variance between courses (arbitrary rule of thumb ICC >.05; see Hayes, 2006). Moreover, for model training, we used ten-fold cross-validation to estimate the prediction error by training and testing the prediction model on different samples of data (Yarkoni & Westfall, 2017). We applied a tune length of 21 to tune the model with respect to the regularization parameters λ and α for the elastic net regression (Helwig, 2017), resulting in 21 combinations of λ and α to identify the combination that best balanced model simplicity and accuracy. To obtain a stable and reliable prediction model, and to quantify the prediction accuracy, we repeated this procedure of cross-validation, training, and testing 100 times. To evaluate the performance of the models, we used explained variance (R^2), the *Root Mean Squared Error* (RMSE), and the *Mean Absolute Error* (MAE) as indices of model fit. To interpret the direction of each predictor variable's effect on learners' engagement with self-tests, we report the respective coefficients for the model with the best tuning parameter λ.

2.3.3. *Variable importance*

To answer the question as to which learner or course characteristics are more suitable for predicting learners' engagement with self-tests, we calculated the average variable importance for all predictor variables in the test sample across all 100 iterations. The values range between 0 and 100, and reflect a transformation of the absolute weights in the linear regression (Grömping, 2009).

2.4. *Sensitivity analyses*

We conducted several sensitivity analyses to check the robustness of our results by contrasting the main model with modified models which varied with respect to methodological approaches, and sample or predictor variable restrictions. For all sensitivity analyses, we compared prediction accuracy and the ten most important predictor variables to examine whether there are differences between models.

2.5. *Statistical software*

All analyses were conducted using *R*-version 4.4.1 (R Core Team, 2025). The *caret* package version 6.0.94 (Kuhn, 2008) was used for the *k*-nearest neighbor imputation and as an interface for modeling and prediction. The *glmnet* package version 4.1.8 (Friedman et al., 2010) was used for elastic net regression.

3. Results

3.1. *Descriptive statistics*

The distribution of the logged percentage of correct test answers (M = .39, SD = .42, range [0; 1]) is characterized by a strong tendency to extremes (see Fig. S2). 36.3 % of the subjects did not answer a single test item correctly.[1] In contrast, 17.2 % of the subjects answered all test questions of a course correctly and thus worked through the course completely. As self-testing was linked to receiving a confirmation of participation, 403 learners (32.0 %) met the threshold required for this confirmation. Among these, ten learners completed a course that required a perfect score (100 %). However, for courses with a lower threshold, achieving the confirmation of participation did not discourage further self-testing, as 98.2 % of learners (386 out of 393) continued using self-tests even after meeting the requirement. Additionally, descriptive statistics and correlations for all scales are summarized in Table S4. A correlation matrix of

[1] This includes participants that a) did not participate in the self-tests, and b) that participated in the self-tests but did not answer any questions correctly.

all 50 predictor variables is displayed in Table E1.

3.2. Elastic net regression

The means and standard deviations of the prediction accuracy (R^2) across 100 random splits of the total sample into training and test sets are displayed in Fig. 1. Overfit of the resulting prediction model is indicated by a difference between training and test sets. Overall, the degree of overfitting was high ($\Delta R^2 = .10$), and 11 % of the variance in the engagement with self-tests could be explained by learner and course characteristics. To further evaluate the prediction model, the average *RMSE* for the train set was .38 compared to .40 for the test set, and the average *MAE* for the train set was .33 compared to .35 for the test set. The best tuning parameters in our elastic net regression were $\alpha = .19$ and $\lambda = .03$. Table S1 shows the coefficients for each predictor variable for the model with the best tuning parameter λ.

3.3. Important variables in predicting the engagement with self-tests

Fig. 2 shows the ten learner and course characteristics with the largest average variable importance for predicting the engagement with self-tests (see Table S1 for the variable importance estimates of all predictor variables). In the following, we will give an overview of the major findings for each data type: First, the most important variable for predicting the engagement with self-tests in terms of process data was the distinction whether a learner participated in our survey once within a single course or multiple times within different courses, which might be an indicator of overall commitment or general learning motivation. Second, in terms of self-report data, learners' enrollment intentions, particularly their intention to receive a confirmation of participation, are especially important for predicting their engagement with self-tests, as they make up four out of ten of the most important variables, thereby highlighting the importance of underlying motivations and external incentives. Lastly, in terms of course characteristics, the course type (i.e., specific area of expertise versus key qualifications/entry-level courses) is the most important variable. However, course type is not consistently among the ten most important variables across sensitivity analyses.

3.4. Sensitivity analyses

Sensitivity analyses were conducted to check whether our findings remain consistent under adjustments to the 1) methodological

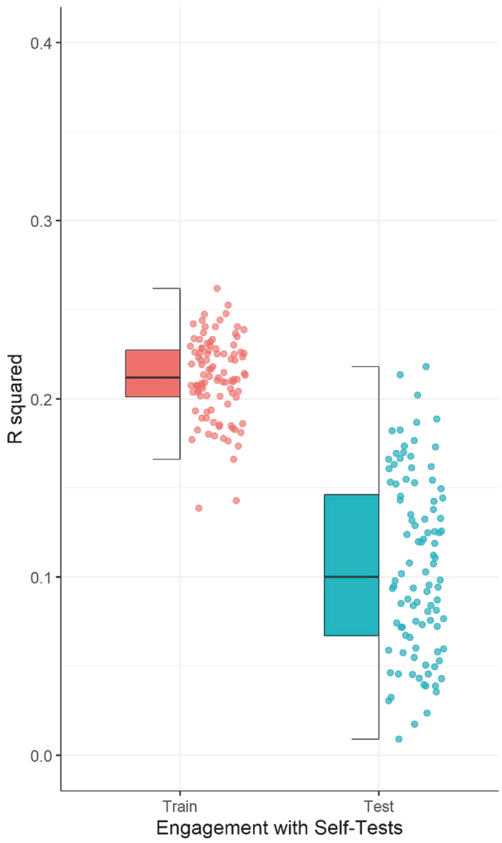

Fig. 1. Accuracy predicting the engagement with self-tests based on learner and course characteristics. $N = 1261$. Left side: The boxplot reflects the interquartile range, the solid line represents the median, and the whiskers represent the minimum/maximum values within 1.5 time the interquartile range. Right side: Jittered point plot of all 100 values for the average R^2.

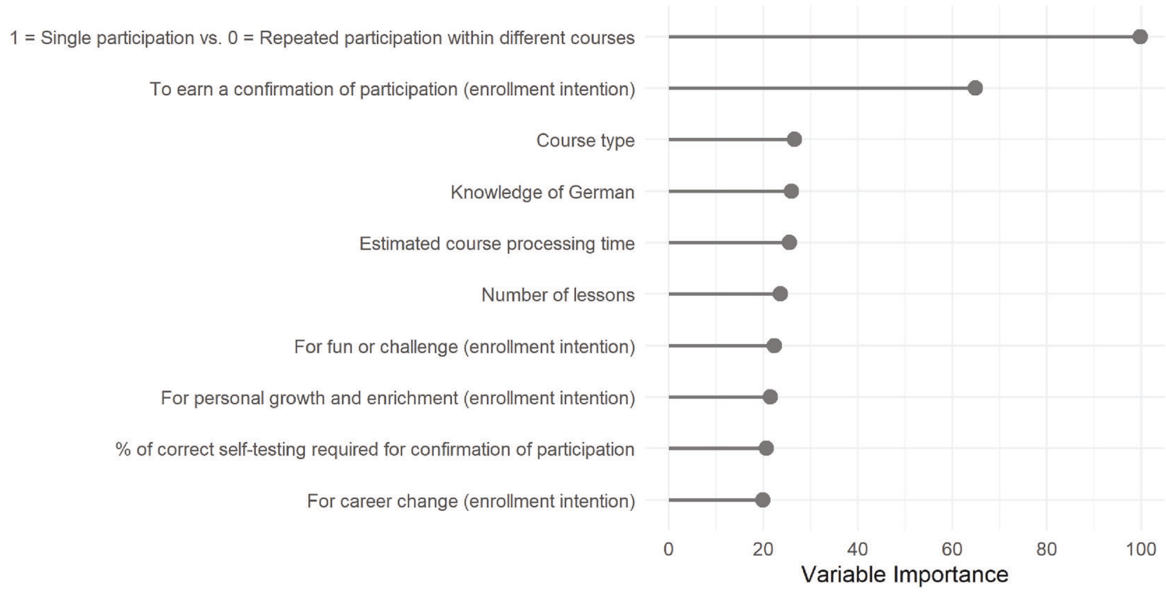

Fig. 2. The ten predictor variables with the largest average variable importance for predicting the engagement with self-tests.

approach, 2) sample composition, and 3) predictor variable selection. Table 1 compares the prediction accuracy of our main model—an elastic net regression of the full analysis sample using blocked resampling—with the prediction accuracy of the modified models. Further, Table S5 summarizes the comparison of the ten most important variables for the prediction of learners' engagement with self-tests across sensitivity analyses.

3.4.1. *Prediction accuracy*

First, we examined the impact of different methodological approaches. To evaluate the impact of resampling approach, we contrasted our main model with an elastic net regression that used simple balanced data splitting based on the outcome instead of blocked resampling (model 1a), which neglects the hierarchical structure of the data. This approach explained more variance and showed less overfitting. In contrast, both, an ordinary multiple linear regression using blocked resampling (model 1b), and a complete case analysis as an alternative approach to handle missing data (Lee et al., 2021; model 1c), explained less variance and resulted in higher overfitting. While neglecting the hierarchical structure of the data might lead to an overestimation of predictive power, the KNN imputation allowed us to retain more information compared to complete case analysis ($N = 1261$ vs. $N = 731$) and yielded plausible model estimates. Overall, the results support the appropriateness of our methodological decisions regarding data splitting, model selection,

Table 1
Sensitivity analyses for the elastic net regression of the full analysis sample using blocked resampling.

		Train set			Test set			ΔTrain-Test		
	N	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	Δ R^2	Δ *RMSE*	Δ *MAE*
Main Model										
Elastic Net Regression (Blocked Resampling)	1261	**.21**	.38	.33	**.11**	.40	.35	**.10**	.02	.02
1) Methodological Approaches										
1a) Elastic Net Regression (Balanced Data Splitting)	1261	**.21**	.38	.33	**.15**	.40	.34	**.06**	.02	.01
Δ Main Model		<.01	<.01	<.01	.04	<.01	−.01	−.04	<.01	−.01
1b) Linear Regression (Blocked Resampling)	1261	**.23**	.37	.32	**.09**	.43	.35	**.14**	.06	.03
Δ Main Model		.02	−.01	−.01	−.02	.03	<.01	.04	.04	.01
1c) Complete Case Analysis	734	**.22**	.38	.34	**.09**	.42	.36	**.13**	.04	.02
Δ Main Model		.01	<.01	.01	−.02	.02	.01	.03	.02	<.01
2) Sample Restrictions										
2a) Unique participation	1205	**.20**	.38	.33	**.09**	.40	.35	**.11**	.02	.02
Δ Main Model		−.01	<.01	<.01	−.02	<.01	<.01	.01	<.01	<.01
2b) At least 1 test completed	803	**.26**	.33	.29	**.12**	.39	.32	**.14**	.06	.03
Δ Main Model		.05	−.05	−.04	.01	−.01	−.03	.04	.04	.01
2c) More tests completed than required for a cop	386	**.28**	.06	.04	**.15**	.07	.05	**.13**	.01	.01
Δ Main Model		.07	−.32	−.29	.04	−.33	−.30	.03	−.01	−.01
3) Predictor Variable Restriction										
Without "intention to obtain a cop"	1261	**.19**	.38	.34	**.09**	.41	.36	**.10**	.03	.02
Δ Main Model		−.02	<.01	.01	−.02	.01	.01	<.01	.01	<.01

Note. ΔTrain-Test = Difference between train and test set. *RMSE* = Root Mean Squared Error. *MAE* = Mean Absolute Error. Other Δs = Difference between main model and modified model. "intention to obtain a cop" = "Intention to obtain a confirmation of participation".

and handling of missing data.

Second, we investigated the effects of different sample compositions by comparing elastic net regressions using blocked resampling for the following subsamples: without cases of multiple participation from the same subject to control for data dependencies (model 2a), containing only learners that have completed at least one test to control for early course dropouts (model 2b), and comprising learners who completed more tests than required for the confirmation of participation (model 2c). Generally, all tested sample restrictions resulted in negligible differences in prediction accuracy as compared to the main model. The most noticeable, albeit small, difference was observed in the subsample of learners who completed more tests than required for the confirmation of participation.

Third, we examined the impact of the variable selection by removing the predictor variable "Intention to earn a confirmation of participation" from the model (model 3) to control for potential masking effects, as completing self-tests were a prerequisite for receiving a confirmation of participation. This also resulted in negligible differences in prediction accuracy as compared to the main model. Taken together, subgroup analyses and the removal of a potential masking predictor variable provided evidence for the robustness of our results.

3.4.2. Important predictors

Important variables for predicting learners' engagement with self-tests are rather similar across models except for those for the subsample of learners who completed more tests than required for the confirmation of participation (model 2c; see Fig. S3). In this case, other predictors were more important compared to those from the main model. However, the resulting different patterns must be interpreted with caution, since the restriction to this highly selective and probably more homogenous subsample leads to severe variance restrictions. When omitting the intention to earn a confirmation of participation as predictor variable, multiple course participation remained by far the most important variable and the other important variables hardly differed from the main model. Hence, the intention to earn a confirmation of participation did not seem to mask effects of the other predictors.

4. Discussion

The rise of online learning opportunities is continuously reshaping the educational landscape, offering learners greater flexibility in choosing when, where, and how to learn (Cha & So, 2020). However, this increase in autonomy also raises the demand for self-regulation (Biwer et al., 2021). Accordingly, learners must take responsibility for their own learning and make informed decisions about which learning techniques to employ (Rivers, 2021). Self-testing can be seen as an effective learning technique (e.g., Carpenter, 2023; Dunlosky et al., 2013; Fiorella & Mayer, 2016). Testing not only allows learners to track their progress, but also promotes active, effortful processing that enhances comprehension and retention. This so-called testing effect is achieved when learners either confirm their understanding by providing correct answers or learn from corrective feedback after making errors (Roediger & Karpicke, 2006). However, while self-testing promotes learning through intensive engagement with the learning material, it also makes learning more effortful (Bjork & Bjork, 2014). Therefore, in a non-formal online learning environment—characterized by its autonomy and lack of external incentives—embracing the challenge of engaging with self-tests until some degree of mastery is achieved seems crucial. To this end, we aimed to predict learners' engagement with self-tests in an authentic learning context. We focused on self-regulated learning using the percentage of correct test answers as a proxy for cognitive engagement. To identify potential predictors, we juxtaposed self-reported learner characteristics (e.g., demographic information, self-regulation, or enrollment intentions), process data from the online survey (e.g., indicators of repetitive response patterns, number of missings within the survey), and course characteristics (e.g., length, content).

4.1. Predicting the engagement with self-tests in non-formal online courses

As expected, not all learners engaged with self-tests to the same extent. While a portion of learners (17.2 %) fully exploited all available self-test opportunities, a considerable share (36.3 %) did not answer a single test item correctly. The remaining 46.5 % of learners varied in their engagement with self-tests, evenly distributed between these two extremes. This finding points to substantial individual differences in learners' engagement with self-tests in non-formal online courses, raising the question of which learner and course characteristics might explain this variability. To tackle this issue, we applied a machine learning approach to identify important variables to predict the engagement with self-tests. Given the context of empirical social science research, the predictive power of our model was acceptable ($R^2 = .11$; Ozili, 2023), offering initial insights and capturing potentially important predictors. However, the large proportion of unexplained variance highlights the complexity of predicting the engagement with self-tests. This is typical for complex human behavior within the learning process, suggesting that there is substantial room for improvement and a need for exploring other predictor variables. At the same time, the most important predictor variables of the present study need to be interpreted with caution given the modest explanatory power of the model.

Nevertheless, the main result of our prediction model was accompanied by two major findings with respect to the applied method: First, even though a R^2 of .11 is on the lower end of the spectrum for explained variance, the elastic net regression still outperforms an ordinary multiple linear regression which justifies the application of the rather complex procedure and highlights the importance of appropriate statistical techniques when examining complex phenomena such as learning behavior. Second, sensitivity analyses emphasized the importance of the implementation of the appropriate data splitting approach, as a non-observance of the hierarchical data structure can quickly lead to biased (i.e., overly-optimistic) prediction models. This is of particular importance in the context of educational research, as nested data structures are usually the case when studying learners from different courses (e.g., students nested in classes, classes nested in schools, and schools nested in school districts; Dedrick et al., 2009).

4.2. Important variables for predicting the engagement with self-tests

Across all main and sensitivity analyses, two predictor variables proved to be particularly important: (1) whether a learner participated in our survey once within a single course or multiple times within different courses, and (2) whether they intended to enroll in order to receive a confirmation of participation. First, participating in multiple courses might indicate that an individual is motivated to learn and committed to the course. These individuals use not only multiple courses as learning opportunities, but they also participate in a study on the same topic multiple times, which might be an indicator for the general motivation to contribute to research as well. This behavior appears to reflect a more serious engagement with the course content overall—including the self-tests. This interpretation aligns with previous findings highlighting the importance of commitment for self-regulated learning in self-paced online learning environments (Kim et al., 2021). Additionally, this finding demonstrates the usefulness of directly observable behavior indicators rather than the use of self-reports to assess learning commitment. It highlights the potential of log data to gain more insight into learning behavior in general (e.g., Klose et al., 2022; Steger et al., 2021). From a practical perspective, fostering commitment to the course (e.g., through gamification elements like badges; Hamari, 2017) might support learners to engage with self-tests as a learning technique (see Rahman et al., 2018 for an overview of frequently used gamification elements). Such elements can help create a more motivating and interactive learning environment that stimulates active participation and enhances potential learning outcomes (Saleem et al., 2022).

Second, the intention to enroll in order to receive a confirmation of participation was a strong predictor of engagement with self-tests. In our study, successful participation in self-tests was necessary to receive such confirmation. Therefore, the offered self-testing opportunities that were intended to have formative value were simultaneously used as a summative judgement of the learner's achievement for course progress (Wiliam & Black, 1996). Interestingly, despite the granted autonomy within non-formal online courses, extrinsic incentives like a confirmation of participation motivate learners to engage with self-tests. Importantly, these incentives did not appear to undermine intrinsic motivation. Most learners continued to engage with self-tests despite reaching the required threshold for a confirmation. This finding is consistent with previous research on non-formal online courses, suggesting that learners' intention to earn a confirmation of participation is associated with more active course engagement, such as watching more videos, participating in discussion forums, and, most importantly, taking more tests (Kizilcec & Schneider, 2015). Therefore, course designers could consider offering a confirmation of participation as a way to promote engagement with self-tests. Perhaps a more fine-grained incentive structure with intermediate learning goals through a badge system could lead to an even more pronounced engagement within the course (Ortega-Arranz et al., 2019) and accordingly to an increase in the engagement with self-tests and enhanced long-term retention and understanding by promoting deeper cognitive processing of the course material.

Apart from these two key predictors, it is particularly remarkable that enrollment intentions in general play a central role for the prediction of the engagement with self-tests. Across models, enrollment intentions consistently make up four out of ten (or even eight out of 20) of the most important variables. This may reflect the diversity of learner goals within non-formal online courses (Moore & Wang, 2021) that have a strong influence on, for example, course completion. Consequently, the longer that learners keep up with the course the greater the likelihood of answering correctly the implemented tests. At the same time, successful learning within non-formal online courses might not look the same for all individuals (Henderikx et al., 2017): Depending on different enrollment intentions and learning goals, individuals might not seek to complete a whole course or master a complete topic, but rather seek very specialized bits of information. Accordingly, "finishing" a course might not be indicative of successful learning behavior for all participants and other learning behavior (such as taking only parts of the course or dropping out early) might be expedient for some individuals (DeBoer et al., 2014). Therefore, future research should consider the enrollment intentions of the learners as a potential explanatory variable within non-formal online courses.

Moreover, regarding course characteristics, we found a consistent pattern of important predictor variables across the different models: estimated course completion time and the total number of lessons consistently contributed to predicting the engagement with self-tests. Therefore, shorter courses seem to be positively linked to learners' engagement with self-tests. Additionally, the threshold for the percentage of correct self-tests required for a confirmation of participation appears to be a relevant factor in predicting engagement. This may also point to the usefulness of including extrinsic incentives in the design of non-formal online courses.

Finally, socio-demographic variables such as age or gender, did not predict the engagement with self-tests. This finding suggests that the ability to embrace challenging learning techniques such as engaging with self-tests is accessible to a diverse range of learners, reinforcing the inclusive potential of non-formal learning for lifelong learning (Werquin, 2010).

4.3. Limitations of the present study and implications for future research

In light of the present results, a few limitations must be mentioned: First, since 89 % of the variance in learners' engagement with self-tests remains unexplained, the question arises as to which other variables might be incrementally predictive. Our examined learner- and course-related characteristics seem not to be sufficient to map out engaging with self-tests in non-formal online courses. One big advantage of online learning environments is the possibility to automatically track information on learning behavior via log data from learners' interactions with the online platform (Gašević et al., 2016). Accordingly, both very general information on the behavioral learning engagement (such as number of logins, or time spent on the course platform), as well as more fine-grained information on which type of learning materials the course participants interact with (for example texts, videos, or discussion boards) might be indicative for the engagement with self-tests (compare with You, 2016 for an overview of the usefulness of log data for the prediction of learning achievement).

Second, in addition to the potential benefits of further studies including log data retrieved from learning platforms, other established psychological constructs might also contribute to a deeper understanding of the engagement with self-tests as a learning technique. The individual perception of the test context might also play an important role. For example with respect to the engagement with self-tests, participants might be driven by performance goals (i.e., either motivated to demonstrate competence, or to avoid showing incompetence) or mastery goals (i.e., seeing the formative assessment as a part of the learning process itself; Elliot & Murayama, 2008). Moreover, learners might associate self-tests with high-stakes summative assessments as a means of evaluating performance instead of an effective learning technique (Dunlosky et al., 2013). This misperception can trigger feelings of (test) anxiety and stress leading to an increase in worry, self-doubt, and interfering thoughts (Weissgerber & Reinhard, 2018), as learners may experience the pressure to perform well, or a fear of failure, as self-tests may expose gaps in their knowledge. This potential negative tendency in the perception of tests is particularly important as learners that are at general risk for these types of learning-related fears might benefit from less structured learning environments that are usually less performance-oriented which might circumvent the negative effects (Bledsoe & Baskin, 2014). In summary, both the investigation of log data as direct indicators of learning behavior, as well as the investigation of other learning-related psychological constructs might contribute further to our understanding of learners' engagement with self-tests and thereby help educators to create learning environments that are beneficial for a larger audience of learners.

Third, future research might also investigate different operationalizations of cognitive engagement with self-tests. The percentage of correct self-tests is only an approximation of such a complex phenomenon. For example, the log data used in the present study cannot distinguish between learners that simply do not engage much with the course content in general (and potentially drop out early), and learners that work through the course and only shy away from the actual engagement with self-tests. This distinction might help to gain more insight into specific characteristics that might influence learners' decision to accept the extra effort that these learning techniques entail. It remains uncertain to what extent our findings can be generalized to other conceptualizations of learning engagement or different test conditions. Other log data—such as the total number of completed tests, or number of test attempts—could provide deeper insights into learners' *behavioral* engagement with self-tests. These information could help to make the distinction more clearly between the quantity of self-testing (i.e., behavioral engagement) and the cognitive effort learners put into mastering these tests (i.e., cognitive engagement). For instance, the total number of test attempts might reveal other patterns, as some learners might answer correctly in a single attempt (e.g., due to prior knowledge), while others might make numerous attempts without answering correctly, yet still participating in self-testing. In this example, the person with multiple test attempts would have shown greater behavioral engagement but still not achieved cognitive engagement.

Fourth, regarding the study design, another potential limitation is the high dropout rate after the initial survey. In the present study, data from a subsequent second and third survey could not be considered in the present analysis, due to large proportions of missing data for these measurement occasions. Especially since our surveys included many motivational variables that are very likely to be related to study dropout, imputation of these variables seemed not a valid option (i.e., missing not at random; MNAR; Donders et al., 2006) This goes along with the observation that our self-report scales at the first measurement occasion did not show much variance, as potentially at the beginning of the course all participants were at least somewhat interested in the course and motivated to start learning. However, with the lack of variance in the initial self-reports, these measures also failed to predict differences in the engagement with self-tests throughout the course. While we additionally aimed to follow-up on participants who dropped out of the study by contacting participants that did not respond to our regularly scheduled surveys after 28 days, the response rate to the follow-up was poor, with only 36 participants (of 1098 contacted individuals) responding. This points to the dilemma that although research within real-life based learning environments is demanded, these setups usually create considerably complex data due to the less controlled environment and more heterogeneous sample. On a more general stance, the increase of ecological validity of these data comes with losses in internal validity since potential confounding variables or certain biases cannot be ruled out. Future study designs might explore new approaches to gain more information, especially on those participants that drop out over time, for example by using more tight-knit sampling designs in the tradition of ecological momentary assessments (Shiffman et al., 2008).

Lastly, the results of the present study might be interpreted with caution, as all analyzed data stem from one specific learning platform. Although the courses varied in certain characteristics such as content and length, they met the same quality standards in terms of course design (e.g., developed by university professors, delivered via the same platform interface). However, unobserved confounders like course reputation, course subject, or perceived difficulty may have influenced learners' engagement with self-tests but were not captured by the presented course characteristics. As not only the participants but also online courses themselves can be very heterogeneous, for example regarding structure or (pedagogical) design (Maya-Jariego et al., 2020), future research might investigate structural differences between courses like course quality, presence of an instructor, time dependent learning schedules, or absence of a confirmation of participation on the engagement with self-tests across different settings. A systematic investigation of course characteristics seems also advisable, as our analyses indicated that the hierarchical structure of the data needs to be taken into account—highlighting that, despite our relatively homogeneous selection of courses, findings cannot be fully generalized from one course to another. Furthermore, the courses also attracted a specific group of participants, resulting in a sample that relies heavily on learners with higher educational backgrounds. Given that the individuals' ability to self-regulate their learning might be confounded with their prior education, it is unclear to what extent the results can be generalized to the general population.

5. Conclusion

In non-formal online courses, a considerable variability in the engagement with self-tests among learners exists, highlighting the importance of learner commitment and extrinsic motivations such as confirmations of participation on this behavior. The application of machine learning models—despite their moderate predictive power—revealed valuable insights and outperformed traditional analytical approaches. Taken together, our results underscored the importance of appropriate analytical techniques for our understanding of complex learning behaviors. Our findings suggested potential practical implications for course designers, including the incorporation of confirmations of participation and more nuanced incentive structures to enhance engagement with self-tests. Course designers might also ensure that course length remains manageable. Additionally, the lack of influence from socio-demographic variables on the engagement with self-tests reinforces the inclusivity of non-formal learning environments, supporting their potential for lifelong learning across various life stages. Regarding the theory on the effectiveness of testing for learning, theoretical models should take into account that there might be entry barriers for learners to accept these challenges, and that the testing environment is crucial because learners need to be committed to the learning opportunity in the first place. Future research should continue to explore additional predictors and refine methods to capture the intricate dynamics of learner behavior, ultimately contributing to the development of more effective and personalized educational experiences. In summary, this study enhances our theoretical understanding of learners' self-regulation by shifting the focus from the effectiveness of self-testing to the question of who engages with self-tests in voluntary settings. Moreover, it offers practical guidance on designing online courses that could encourage the wider use of self-testing as a learning technique, such as the implementation of extrinsic incentives like confirmations of participation. Overall, the study advances our understanding of learner engagement in non-formal online courses and provides concrete recommendations for course design to encourage the use of self-tests.

CRediT authorship contribution statement

Maria Klose: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Philipp Handschuh:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation, Conceptualization. **Diana Steger:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Cordula Artelt:** Writing – review & editing, Supervision, Conceptualization.

Declarations of competing interest

None. Supplemental materials, syntax and data are available online at the Center for Open Science, 2025 (https://osf.io/5za3d/). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The data used in the present study partly overlaps with the data published in another study (Klose et al., 2024), which leads to some similarities within the method section.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2025.105507.

Data availability

Data, syntax, and all supplemental materials are available online within the Open Science Framework (Center for Open Science, 2025): https://osf.io/5za3d/.

References

Abbakumov, D., Desmet, P., & Van Den Noortgate, W. (2020). Psychometrics of MOOCs: Measuring learners' proficiency. *Psychologica Belgica, 60*(1), 115–131. https://doi.org/10.5334/pb.515

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review, 33*, 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369–386. https://doi.org/10.1002/pits.20303

Bertilsson, F., Stenlund, T., Sundström, A., & Jonsson, B. (2024). Self-regulated use of retrieval practice: Associations with individual differences in non-cognitive and cognitive factors. *European Journal of Psychology of Education, 39*(4), 4091–4111. https://doi.org/10.1007/s10212-024-00845-2

Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning and Teaching, 20*(1), 21–39. https://doi.org/10.1177/1475725720973494

Bilger, F., & Strauß, A. (2022). *Weiterbildungsverhalten in Deutschland 2020: Ergebnisse des Adult Education Survey—AES-Trendbericht [Continuing education behavior in Germany 2020: Results of the Adult Education Survey—AES Trend Report]*. Federal Ministry of Education and Research. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/1/31690_AES-Trendbericht_2020.html.

Biwer, F., Wiradhany, W., Oude Egbrink, M., Hospers, H., Wasenitz, S., Jansen, W., & De Bruin, A. (2021). Changes and adaptations: How university students self-regulate their online learning during the COVID-19 pandemic. *Frontiers in Psychology, 12*, Article 642593. https://doi.org/10.3389/fpsyg.2021.642593

Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59–68). Worth Publishing.

Bledsoe, T. S., & Baskin, J. J. (2014). Recognizing student fear: The elephant in the classroom. *College Teaching, 62*(1), 32–41. https://doi.org/10.1080/87567555.2013.831022

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* McGraw-Hill.

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education, 17*(2). https://doi.org/10.1186/s41239-019-0176-8

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Carpenter, S. K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review, 35*(4), 96. https://doi.org/10.1007/s10648-023-09811-8

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology, 1*, 496–511. https://doi.org/10.1038/s44159-022-00089-1

Center for Open Science. (2025). *Open Science Framework*. https://osf.io/.

Cha, H., & So, H.-J. (2020). Integration of formal, non-formal and informal learning through MOOCs. In D. Burgos (Ed.), *Radical solutions and open science: An open approach to boost higher education* (pp. 135–158). Singapore: Springer. https://doi.org/10.1007/978-981-15-4276-3_9.

De Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review, 35*(2), 41. https://doi.org/10.1007/s10648-023-09766-w

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course": Reconceptualizing educational variables for massive open online courses. *Educational Researcher, 43*(2), 74–84. https://doi.org/10.3102/0013189X14523038

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*(1), 69–102. https://doi.org/10.3102/0034654308325581

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice, 55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In J. Dolin, & R. Evans (Eds.), *Transforming assessment* (Vol. 4, pp. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3_3.

Donders, A. R. T., Van Der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Eriksson, T., Adawi, T., & Stöhr, C. (2017). "time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education, 29*, 133–146. https://doi.org/10.1007/s12528-016-9127-8

Fellman, D., Lincke, A., & Jonsson, B. (2020). Do individual differences in cognition and personality predict retrieval practice activities on MOOCs? *Frontiers in Psychology, 11*, 2076. https://doi.org/10.3389/fpsyg.2020.02076

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Glaser, J., & Richter, T. (2023). The testing effect in the lecture hall: Does it depend on learner prerequisites? *Psychology Learning and Teaching, 22*(2), 159–178. https://doi.org/10.1177/14757257221136660

Goopio, J., & Cheung, C. (2021). The MOOC dropout phenomenon and retention strategies. *Journal of Teaching in Travel & Tourism, 21*(2), 177–197. https://doi.org/10.1080/15313220.2020.1809050

Greenwood, C. J., Youssef, G. J., Letcher, P., Macdonald, J. A., Hagg, L. J., Sanson, A., Mcintosh, J., Hutchinson, D. M., Toumbourou, J. W., Fuller-Tyszkiewicz, M., & Olsson, C. A. (2020). A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One, 15*(11), Article e0242730. https://doi.org/10.1371/journal.pone.0242730

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician, 63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational Motivation Scale (SIMS). *Motivation and Emotion, 24*(3), 175–213. https://doi.org/10.1023/A:1005614228250

Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior, 71*, 469–478. https://doi.org/10.1016/j.chb.2015.03.036

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32*(4), 385–410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Helwig, N. E. (2017). Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology, 13*(1), 1–19. https://doi.org/10.20982/tqmp.13.1.p001

Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention–behavior gap. *Distance Education, 38*(3), 353–368. https://doi.org/10.1080/01587919.2017.1369006

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 42*(1), 80–86. https://doi.org/10.1080/00401706.2000.10485983

Jaldemark, J., Håkansson Lindqvist, M., Mozelius, P., & Ryberg, T. (2021). Editorial introduction: Lifelong learning in the digital era. *British Journal of Educational Technology, 52*(4), 1576–1579. https://doi.org/10.1111/bjet.13128

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology, 113*(5), 972–985. https://doi.org/10.1037/edu0000627

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 487–514). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9.

Kim, D., Jung, E., Yoon, M., Chang, Y., Park, S., Kim, D., & Demir, F. (2021). Exploring the structural relationships between course design factors, learner commitment, self-directed learning, and intentions for further learning in a self-paced MOOC. *Computers & Education, 166*, Article 104171. https://doi.org/10.1016/j.compedu.2021.104171

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology, 115*, Article 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

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. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction, 22*(2), 1–24. https://doi.org/10.1145/2699735

Klingsieck, K. B. (2018). Kurz und knapp – die Kurzskala des Fragebogens „Lernstrategien im Studium" (LIST) [Short and sweet – the short version of the questionnaire "Learning Strategies of University Students" (LIST)]. *Zeitschrift für Padagogische Psychologie, 32*(4), 249–259. https://doi.org/10.1024/1010-0652/a000230

Klose, M., Handschuh, P., Steger, D., & Artelt, C. (2024). Not easy to get off track: Motivational trajectories of learners completing a non-formal online course. *Computers in Human Behavior, 159*, 108322. https://doi.org/10.1016/j.chb.2024.108322

Klose, M., Steger, D., Fick, J., & Artelt, C. (2022). Decrypting log data: A meta-analysis on general online activity and learning outcome within digital learning environments. *Zeitschrift für Psychologie, 230*(1), 3–15. https://doi.org/10.1027/2151-2604/a000484
Kruppe, T., & Baumann, M. (2019). *Weiterbildungsbeteiligung, formale Qualifikation, Kompetenzausstattung und Persönlichkeitsmerkmale [Participation in continuing training, formal qualifications, skills and personality traits]* (Research Report No. 1/2019. IAB-Forschungsbericht). Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg. http://hdl.handle.net/10419/204767.
Kuhn, M. (2008). Building predictive models in *R* using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05
Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research, 83*(3), 432–479. https://doi.org/10.3102/0034654313480891
Lee, K. J., Tilling, K. M., Cornish, R. P., Little, R. J. A., Bell, M. L., Goetghebeur, E., Hogan, J. W., & Carpenter, J. R. (2021). Framework for the treatment and reporting of missing data in observational studies: The treatment and reporting of missing data in observational studies framework. *Journal of Clinical Epidemiology, 134*, 79–88. https://doi.org/10.1016/j.jclinepi.2021.01.008
Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70*(4), 647–671. https://doi.org/10.1177/0013164409355699
Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199. https://doi.org/10.3758/BF03194051
Maya-Jariego, I., Holgado, D., González-Tinoco, E., Castaño-Muñoz, J., & Punie, Y. (2020). Typology of motivation and learning intentions of users in MOOCs: The MOOCKNOWLEDGE study. *Educational Technology Research & Development, 68*, 203–224. https://doi.org/10.1007/s11423-019-09682-3
Moore, R. L., & Wang, C. (2021). Influence of learner motivational dispositions on MOOC completion. *Journal of Computing in Higher Education, 33*, 121–134. https://doi.org/10.1007/s12528-020-09258-8
Mozelius, P., Cleveland-Innes, M., Hakansson Lindqvist, M., & Jaldemark, J. (2024). Critical aspects of a higher education reform for continuous lifelong learning opportunities in a digital era. *Electronic Journal of Knowledge Management, 22*(1), 26–39. https://doi.org/10.34190/ejkm.22.1.3258
Ortega-Arranz, A., Er, E., Martínez-Monés, A., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., & Muñoz-Cristóbal, J. A. (2019). Understanding student behavior and perceptions toward earning badges in a gamified MOOC. *Universal Access in the Information Society, 18*(3), 533–549. https://doi.org/10.1007/s10209-019-00677-8
Ozili, P. K. (2023). The acceptable R-square in empirical modelling for social science research. In C. A. Saliya (Ed.), *Social research methodology and publishing results: A guide to non-native English speakers* (pp. 134–143). IGI Global. https://doi.org/10.4018/978-1-6684-6859-3.ch009.
Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science, 6*(3), Article 25152459231162559. https://doi.org/10.1177/25152459231162559
Pastötter, B., & Bäuml, K.-H. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging, 34*(2), 242–250. https://doi.org/10.1037/pag0000307
Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3.
R Core Team. (2025). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing [Computer software] https://www.R-project.org/.
Rahman, M. H. A., Panessai, I. Y., Noor, N. A. Z. B. M., & Salleh, N. S. M. (2018). Gamification elements and their impacts on teaching and learning – A review. *The International journal of Multimedia & Its Applications, 10*(6), 37–46. https://doi.org/10.5121/ijma.2018.10604
Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review, 33*(3), 823–862. https://doi.org/10.1007/s10648-020-09578-2
Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881
Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x
Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre, & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 1–36). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00001-6.
Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559
Saleem, A. N., Noori, N. M., & Ozdamli, F. (2022). Gamification applications in e-learning: A literature review. *Technology, Knowledge and Learning, 27*, 139–159. https://doi.org/10.1007/s10758-020-09487-x
Schwier, R. A., & Seaton, J. X. (2013). A comparison of participation patterns in selected formal, non-formal, and informal online learning environments. *Canadian Journal of Learning and Technology, 39*(1). https://www.learntechlib.org/p/54424/.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415
Steger, D., Schroeders, U., & Wilhelm, O. (2021). Caught in the act: Predicting cheating in unproctored knowledge assessment. *Assessment, 28*(3), 1004–1017. https://doi.org/10.1177/1073191120914970
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B, 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics, 17*(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520
Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One, 14*(11), Article e0224365. https://doi.org/10.1371/journal.pone.0224365
Virtuelle Hochschule Bayern. (2025). *OPEN Vhb*. https://open.vhb.org/?lang=en.
Wang, Q., Hall, G. J., Zhang, Q., & Comella, S. (2024). Predicting implementation of response to intervention in math using elastic net logistic regression. *Frontiers in Psychology, 15*, Article 1410396. https://doi.org/10.3389/fpsyg.2024.1410396
Weissgerber, S. C., & Reinhard, M.-A. (2018). Pilot study on the relationship of test anxiety to utilizing self-testing in self-regulated learning. *International Journal of Psychological Studies, 10*(4), 95. https://doi.org/10.5539/ijps.v10n4p95
Werquin, P. (Ed.). (2010). *Recognising non-formal and informal learning: Outcomes, policies and practices.* OECD Publishing.
Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537–548. https://doi.org/10.1080/0141192960220502
Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309
Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393
You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education, 29*, 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003
Zepeda, C. D., Martin, R. S., & Butler, A. C. (2020). Motivational strategies to engage learners in desirable difficulties. *Journal of Applied Research in Memory and Cognition, 9*(4), 468–474. https://doi.org/10.1016/j.jarmac.2020.08.007
Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society - Series B: Statistical Methodology, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x