Education and Information Technologies (2025) 30:10979–10995
https://doi.org/10.1007/s10639-024-13261-2

# Conventional teaching vs. e-learning: A case study of German undergraduate biology students

**Tim Bauermeister**[1] · **Nina Janßen**[1] · **John-Oliver Engler**[2,3]

Received: 1 August 2024 / Accepted: 5 December 2024 / Published online: 24 December 2024
© The Author(s) 2024

## Abstract

Digital learning environments such as virtual laboratories have been proposed to augment or even replace conventional forms of teaching in recent years. While there is some evidence on comparable learning outcomes between real and virtual labs, empirical evidence on the effectiveness of virtual labs as compared to conventional face-to-face teaching is largely lacking. Here, we used the virtual lab platform Labster in a first-year undergraduate course on general biology to evaluate its effectiveness against conventional teaching using a repeated measure randomized control trial (treatment group n = 131, control group n = 108). We find no evidence for a difference in terms of learning success between treatment and control groups, consistent with empirical studies on real vs. virtual labs. However, Labster exposure led to a more positive attitude towards e-learning, and a positive user experience was the strongest predictor of individual learning success, highlighting the importance of engagement in digital learning environments.

**Keywords** Virtual laboratories · Learning outcomes comparison · Undergraduate biology education · User experience in digital learning · Attitudes towards e-learning

✉ Tim Bauermeister
tim.bauermeister@uni-vechta.de

[1] Faculty II, Biology, University of Vechta, Driverstraße 22, Vechta 49377, Germany

[2] Vechta Institute of Sustainability Transformations in Rural Areas, University of Vechta, Driverstraße 22, 49377 Vechta, Germany

[3] Center for Methods, Leuphana University Lüneburg, Universitätsallee 1, Lüneburg 21335, Germany

## 1 Introduction

Digitalization has significantly and continuously transformed the educational landscape over the past decades through the emergence of new teaching media and instructional methods. This development was further accelerated by the COVID-19 pandemic and the resulting shift to enforced distance learning and e-learning (Ray & Srivastava, 2020). As the SAMR model by Puentedura (2006) demonstrates, educational technology can be integrated into teaching at four distinct levels: as a direct substitution for analog options, as an augmentation with functional enhancement, for the significant modification of existing tasks, or for a complete redefinition and the creation of entirely new tasks. However, as described by Herzig (2014), the use of digital media does not necessarily lead to improved learning outcomes. Moreover, generalizable conclusions about the effects of specific media in the classroom cannot be made; such assessments are always context-dependent (Herzig, 2014). This underscores the need for practical educational research to examine the use of selected media within context-specific case studies.

From the vast pool of digital media, virtual labs such as Labster are a learning tool of particular interest in biology, as they provide a virtual representation of specific scientific methodologies. The significance of virtual labs is highlighted by the bibliometric analysis of Heradio et al. (2016), who concluded that the number of papers on virtual or remote labs has multiplied between 1993 and 2014. While Heradio et al. (2016, 32) refer to "more than four thousand articles" for their study period, a query in Web of Science using the search term "virtual laboratories" reveals that more than 60,000 peer-reviewed English journal articles, more than 27,000 of which are Open Access, are available as of November 2024. Since 2016, a four-digit number of articles has been published annually under this keyword, with more than 3,000 articles published each year between 2021 and 2023. These figures underscore the continued relevance of this research area.

Studies have already demonstrated a variety of educational benefits of virtual labs. Makransky et al. (2016), Thisgaard and Makransky (2017), Coleman and Smith (2019), Bonde et al. (2014) and De Vries and May (2019) report positive effects of digital labs on student interest and motivation. Moreover, Makransky et al. (2016), Thisgaard and Makransky (2017), Schäfers et al. (2020) and Kolil et al. (2020) highlight positive influences on self-efficacy. Studies have also been carried out in the area of learning outcomes that indicate the potential of virtual laboratories as Darrah et al. (2014) and Hawkins and Phelps (2013) conclude that virtual laboratories can achieve equivalent learning outcomes in comparison to real laboratories.

Additionally, recent comparative studies have examined the use of virtual laboratories. Byukusenge et al. (2022) reviewed 26 studies on the application of virtual labs in biology and concluded that these labs can be effective in various areas, such as enhancing conceptual understanding or serving as preparation for hands-on laboratory work. However, they caution that the effectiveness of virtual labs “may be greatly influenced by how they are used in the classroom” (Byukusenge et al., 2022, 12), highlighting the need for further content-specific research. This finding aligns with Herzig’s earlier study (2014), which suggests that the

success of digital methods is not guaranteed and must be evaluated on a case-by-case basis. Another comparative study, this time in the field of chemistry, was conducted by Chan et al. (2021), who reviewed 76 papers. While the overall tone of the findings is positive, indicating that virtual labs can serve as an effective complementary tool or even as an alternative to hands-on laboratories, a notable aspect is that several of the reviewed studies question the viability of using virtual labs as a full replacement for real laboratory experiences.

On a structural level, a primary reason for utilizing virtual labs is their independence from physical lab equipment and facilities. Consequently, virtual labs can serve as a cost-effective alternative to both hands-on and remote labs (Chatterjee, 2021; Diwakar et al., 2012; Glassey & Magalhães, 2020). They provide learners with the opportunity to engage with course material online, from any location and at their own pace (Bhargava et al., 2006; Waldrop, 2013).

This insight into the state of research reveals that a significant proportion of published studies on virtual laboratories are based on comparisons with real laboratories. If we categorize this approach using the aforementioned SAMR model by Puentedura (2006), it becomes evident that substituting real lab experiences with virtual labs aligns with the first levels of the model (mostly Substitution and Augmentation, sometimes also Modification). Thus, there is a shift from the physical lab space to a virtual environment; however, the tasks and topics covered remain relatively similar.

This study, however, takes a different approach: In our specific context of a course for first-year biology students at the University of Vechta, the virtual lab Labster offers the opportunity to adopt an entirely new seminar concept. Due to insufficient lab facilities, it is not feasible to conduct the course in a real laboratory. Consequently, for many topics, lecturers have so far relied on traditional teaching methods without any experimental components.

But the comparison of virtual labs to traditional face-to-face teaching methods seems largely unexplored. Studies by Bonde et al. (2014) and Makransky et al. (2016) in the context of medicine and biotechnology showed that learning outcomes were improved through the use of the virtual lab platform Labster compared to conventional teaching methods. These findings suggest the potential of virtual labs. However, since they refer to specific contexts and topics, and the amount of publications on the comparison of virtual labs to traditional teaching methods is quite sparse in general, it remains unclear whether the results can be transferred to our specific context.

For this study, we conducted a pilot study over one semester in an undergraduate biology course at a small university in northwest Germany (University of Vechta) using a virtual lab by Denmark-based e-learning company Labster.[1] Labster is a virtual platform that offers interactive simulations of lab experiments in various scientific disciplines, including natural science subjects such as biology. These virtual experiments allow learners to gain experience by hypothesizing, experimenting, collecting and analyzing data, and exploring scientific concepts and methods in a practice-based way. Labster covers a variety of topics in biology, including cell and

[1] www.labster.com

molecular biology, genetics, human biology, evolution and ecology. These focus areas illustrate that the simulations address complex topics that could normally only be explored in a physical laboratory.

The aim of the pilot phase was to use a comparative research design to determine how course participants who use Labster differ in terms of knowledge gain from course participants who are taught the same topic through a traditional instruction method as is usually used in our university teaching. The results and observations are presented in this paper. The paper proceeds as follows: Sect. 2 describes our experimental design, data gathering and methods. Section 3 lists and describes the results before Sect. 4 discusses them in the context of existing related literature. Section 5 draws conclusions.

## 2 Methods and data

### 2.1 Experimental design

Labster was tested as part of a seminar accompanying the lecture "general biology". Both the lecture and the seminar are mandatory for all undergraduate biology students at the University of Vechta. On the one hand, the seminar serves to deepen the content taught in the lecture, but on the other hand it also serves to practice biological and scientific working methods. The recommendation for this module is to study it in the first semester.

Due to the large number of participants, the seminar was split into four identical courses (A, B, C and D). Two courses (B and C) worked with Labster on four of the total 14 course days, the other two courses served as a control group. Our students were randomly assigned to one of the four courses without knowing beforehand whether the chosen course would be part of the experimental group or the control group. All subjects, or their legal guardian, gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki.[2] We illustrate our experimental design in Fig. 1.

The topics taught using Labster simulations were: Diffusion and osmosis, introduction to microbiology (gram stain), phosphorus cycle, eutrophication of a water body. We also attempted to use Labster on the topic of introduction to microscopy but this had to be canceled on short notice because the internet connection in the course room turned out to be insufficient. As a result, further testing took place in a computer room at a nearby school on two course days. The participants could decide for themselves whether they wanted to complete the simulation on an iPad provided or on their private devices. On the other two dates Labster was provided as a compulsory task while distance learning sessions took place.

We taught the students in the control group the corresponding seminar topics in a different, comparable way in terms of content. Traditional teaching methods encompassed video demonstrations in a physical or virtual classroom setting,

[2] German law, currently or at the time this study was conducted, does not require ethics approval statements in cases of non-medical research involving human subjects.

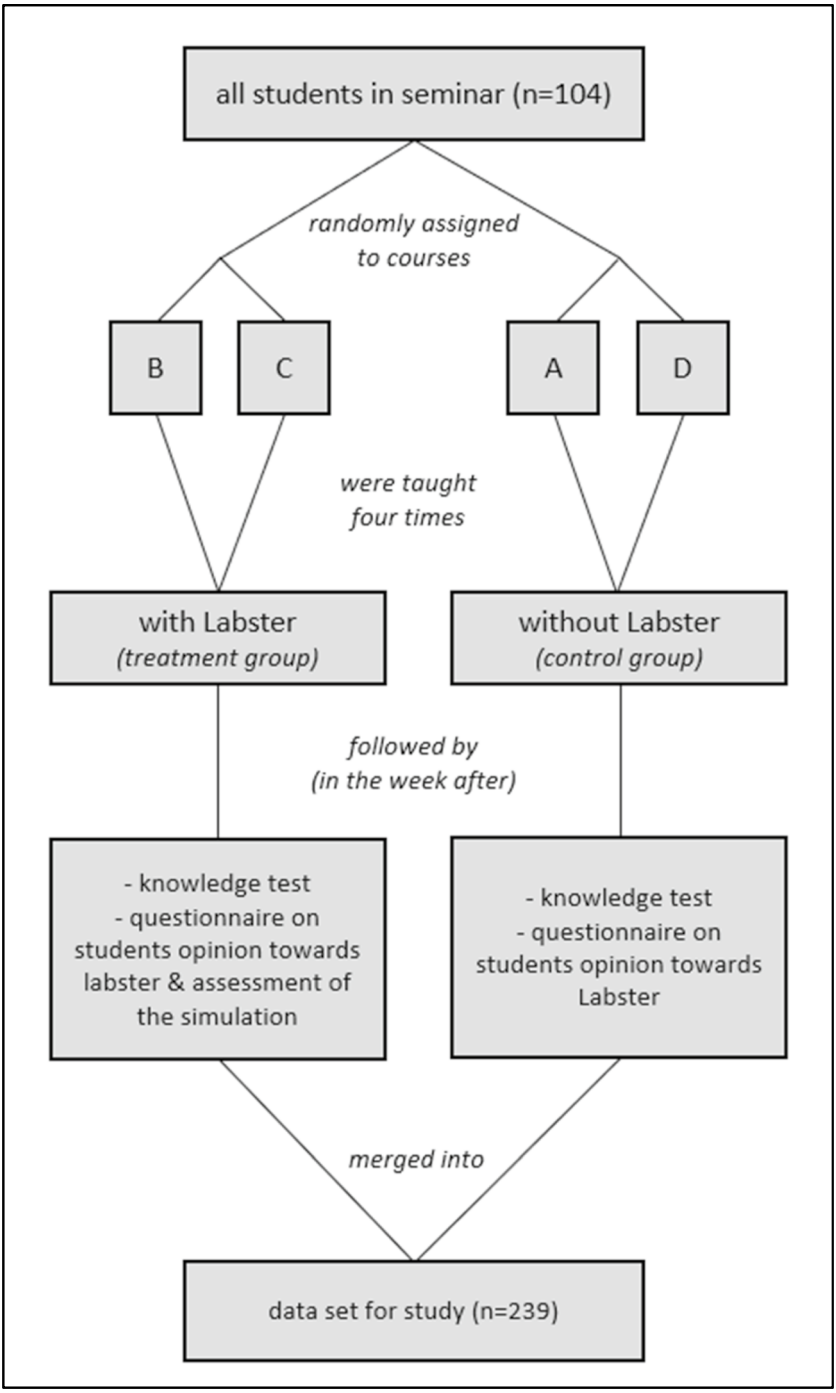


**Fig. 1** Schematic overview of our repeated measures experimental design with one treatment and one control group

didactic lectures, group work, and self-directed internet searches to address practice tasks. Care was taken to ensure that the amount of information and the chosen content were similar to the chosen Labster simulations. For this purpose, we played the Labster simulation before designing the seminar day and reviewed the exercise questions as well as the background texts available in the simulation.

The students in the experimental group had 20 to 40 min to complete the simulations, depending on the unit. Attention was paid to ensure that there was no time pressure. We gave a short introduction to the respective topic before the students started the simulation without giving away essential learning contents of the simulation. The students worked individually and followed the tasks provided by Labster for each respective simulation. These tasks primarily consisted of quiz questions as well as activities to be completed in the virtual laboratory, such as donning safety goggles and gloves as a preparation for workplace safety. In Labster simulations, progress is linked to providing correct answers or performing the correct actions; when students provide an incorrect solution, they are typically placed into a repetition loop. To assist in finding the correct answer, Labster offers background texts that further elaborate on the topic and can be optionally reviewed by the students. Additionally, no supplementary materials on the topics were distributed during the session, though students were encouraged to approach us if they had difficulty understanding content provided by Labster.

Every time Labster was used, the participants from both the experimental group and the control group took part in a survey one week after. The survey consisted of two parts. Part 1 was a knowledge test consisting of five questions (mixture of multiple choice, single choice, free answers) on the content of the previous week, which was taught with Labster or analogue. Between 0 and 5 points could be achieved. Part 2 was a questionnaire for the students in the experimental group to give their assessment of the simulation played. Both groups also provided information on their opinion towards Labster and whether they would like to work with Labster in the future.

For the purpose of evaluation, all surveys were merged into one data set. Provided that the students were present on all four days, they can appear up to four times in the data set. The sample (treatment group $n = 131$, control group $n = 108$) therefore provides information about the number of completed questionnaires and not about the number of different persons in the data set. Through an anonymised personal identifier, it was possible to identify in the data set how often a person participated in the surveys.

### 2.2 Variables and data preparation

The variables mainly used in our study were demographic variables (age, gender, semester), the results of the post-course knowledge test and two variables constructed from the questionnaire on students' opinions and perceptions of Labster. We provide translated versions of our questionnaire in Table 1. We assessed

**Table 1** Overview on items and questions of the post-course questionnaire

| Item—question | Question |
|---|---|
| 1–1 | I would like to work with Labster more often during my studies |
| 1–2 | I would like to work with Labster more often in the course of my studies, if that means that I get to engage with content that could not otherwise be covered at University of Vechta |
| 2 | How did you like the Labster unit on […] in the categories mentioned? |
| 2–1 | Overall impression? |
| 2–2 | Visual representation? |
| 2–3 | Text design (background information, tasks, dialogs)? |
| 2–4 | Scenario / Story? |
| 2–5 | Duration of the unit? |
| 2–6 | Amount of information? |

students' openness to e-learning formats such as Labster with two questions, one asking about their preferences towards an increased use of Labster during their studies, and one asking the same thing, but with an additional clause that highlighted that using Labster would mean learning about subjects that were otherwise impossible to teach, for example because of insufficient laboratory facilities (cf. 1–1 and 1–2 in Table 1). We assessed students' user experience with a set of six questions about different aspects of their respective experiences during their Labster sessions (cf. 2–1 through 2–6 in Table 1).

For each question, students had to answer on a 6-point Likert scale from 1 ("strongly agree" / "very good") to 6 ("strongly disagree" / "unsatisfactory"), where answers to the questions for item 2 followed the grading scheme used in the German school system (1 = very good 2 = good, 3 = satisfactory, 4 = sufficient, 5 = poor, 6 = unsatisfactory). We reverse coded the answers prior to further analysis to ensure results were easier to interpret. In particular, this means that larger scores on any of our questions signified a more positive attitude or experience. This procedure also explains the min (max) scores of 2 (12) on the variable openness to e-learning and 6 (36) on the variable user experience (cf. Table 2).

**Table 2** Descriptive statistics of the data set

| Variable | Overall (median, min, max),$N = 239$ | Labster / treatment group (median, min, max),$N = 131$ | Conventional / control group (median, min, max),$N = 108$ |
|---|---|---|---|
| Age | 20.7 (21, 19, 35) | 20.5 (20, 19, 27) | 21.0 (21, 19, 35) |
| Gender | 1\|36\|200 (d\|m\|f) | 1\|15\|115 (d\|m\|f) | 0\|21\|85 (d\|m\|f) |
| Semester | 1.16 (1, 1, 7) | 1.13 (1, 1, 7) | 1.19 (1, 1, 3) |
| Openness to e-learning | 8.48 (9, 2, 12) | 8.93 (9, 3, 12) | 7.86 (8, 2, 12) |
| User experience | 26.68 (28, 9, 36) | 26.68 (28, 9, 36) | NA |

### 2.3 Analyses

For comparison and analysis of learning success in the two methods investigated, we calculate and report Welch's t-test, the Bayes factor and a one-way ANOVA. Moreover, we model learning success as measured by the points achieved in a 5-question post-course test for each subject matter as a function of four independent variables in a statistical candidate modeling approach following the information-theoretic paradigm of statistical inference (Burnham & Anderson, 2002). In particular, we use Akaike's Information Criterion for small samples (AICc) to rank all candidate models relative to each other. Independent variables include respondents' gender (G), age (A), openness to e-learning (O) and their rating of the user experience when using Labster (U). Including the empty set (i.e. the null model), it is possible to combine these independent variables into 24 = 16 different additive combinations of predictors. Because we do not see any strong scientific reasons for excluding specific combinations of predictors a priori, the candidate model set therefore includes 16 generalized linear regression models, each with a log link, a negative binomial error structure, and 'points achieved' as the dependent variable (see Sect. 3.3.3). All candidate models are additive, i.e. we do not consider any interaction effects between independent variables. In particular, our analysis aims at analyzing the success factors of an e-learning format in university undergraduate biology education. We used version 4.2.2 of the R programming language (R Core Team, 2022) to perform our analyses.

## 3 Results

### 3.1 Descriptive statistics

Table 2 lists some descriptive statistics of the data set.

As expected, most participants are in the first semester of their undergraduate studies in biology, which is also reflected in the variable age. Because the students were assigned randomly to the four courses, there is no self-selection bias. Hence, we have the incidental finding that treatment and control group differed significantly in their openness to e-learning, both according to Welch's t-test ($t(\mathrm{df} = 183.79) = 4.186, p < .001$) and according to the Bayes factor ($\mathrm{BF}_{10} = 591.15$).

### 3.2 Learning success

First, we tested for differences in mean scores attained in the post-lab quiz between control and treatment group. A Welch two-sample $t$-test finds no evidence to reject the null hypothesis of equal means ($t(\mathrm{df} = 230.87) = -1.241, p = .215$). The Bayes factor indicates that the difference in group means is indeed zero ($\mathrm{BF}_{10} = 0.293$). Thus, there is no evidence that the two groups differ systematically in terms of learning success.

A one-way ANOVA revealed that there was no statistically significant difference in post-course test scores between at least two groups ($F(3,235) = 0.58, p = 0.63$). The right-hand side of Fig. 2 shows the associated boxplots.

As to differences in learning success according to subject taught, the only non-zero difference occurred with phosphorus cycle (Fig. 3). Due to strongly unbalanced data (49:32, 7:16, 45:36 and 30:24, Labster:Conventional data points, from left to right in Fig. 3), calculation of a MANOVA was not possible. A Welch two-sample $t$-test suggests rejection of the null hypothesis of equal means ($t(\mathrm{df} = 51.587) = -2.878, p = .006$). The Bayes factor further indicates that the

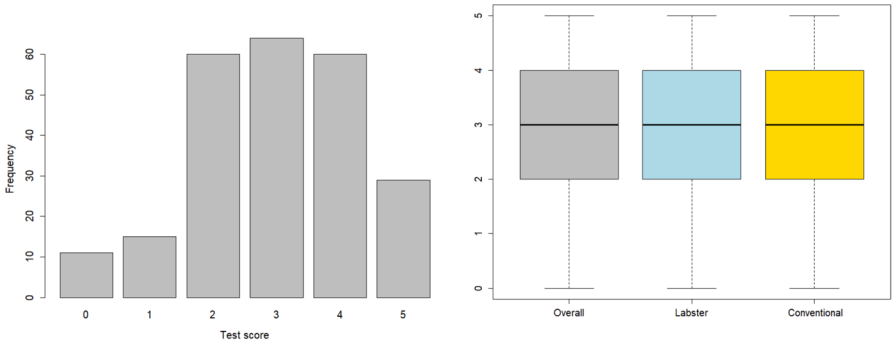


**Fig. 2** Barplot of the overall sample test score distribution (left panel) and boxplots of the scores achieved overall and by mode of teaching (right panel). Based on various statistical criteria, there is no evidence that learning success differed systematically by mode of teaching

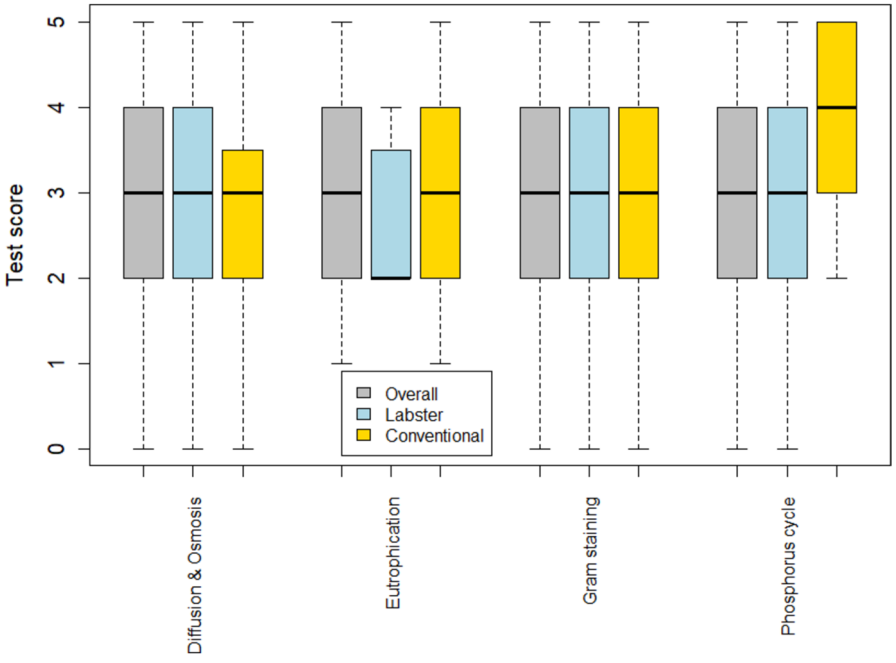


**Fig. 3** Boxplots of the test scores achieved in the post-course knowledge test per subject taught, sample and mode of teaching

difference in group means is indeed non-zero ($BF_{10} = 5.97$). Thus, there is good evidence that learning success was on average about one full point greater in the conventional group than in the Labster group for the course subject of phosphorus cycle. For all other settings, there is no evidence for differing learning success (cf. Fig. 3).

### 3.3 What makes (or breaks) a successful Labster unit?

Candidate modeling indicates three models as best in the candidate set, i.e. within two AICc units of the AICc-best model (Table 3), which is commonly recommended as a cutoff value for considerable support (Burnham & Anderson, 2002). These three models, U, O + U and NULL, have a cumulative Akaike weight of 52.5%, roughly half of which (27.7%) can be attributed to the single best model U. Pseudo $R^2$ measures indicate that data variability explained by the models was relatively modest throughout the entire candidate set, with 0.9% and 1.5% for the winning models and 3.6% for the maximal model G + A + O + U. However, the maximal model's Akaike weight was only 0.9% and it was not within two units of the best model as measured by AICc. Except for the NULL model, which is the intercept-only model, all models within 2 units of the AICc-best model contained the variable user experience (U). The 95% confidence set, which amounts to the smallest possible set of those models closest to the best model for which cumulative Akaike weight surpasses 95%, contains 12 out of 16 models.

**Table 3** Results of the candidate modeling with the candidate set of 16 possible models built from the set of four independent variables

| Model | No. parameters | AICc | Δ AICc | Akaike weight [%] | Log-likelihood | Cum. Akaike weight [%] | Nagelkerke's $R^2$ [%] |
|---|---|---|---|---|---|---|---|
| U | 3 | 464.04 | 0 | 27.69 | −228.93 | 27.69 | 0.94 |
| NULL | 2 | 465.51 | 1.47 | 13.27 | −230.71 | 40.97 | 0.00 |
| O + U | 4 | 465.80 | 1.76 | 11.48 | −228.74 | 52.47 | 1.47 |
| A + U | 4 | 466.14 | 2.10 | 9.70 | −228.91 | 62.15 | 0.98 |
| O | 3 | 466.59 | 2.55 | 7.75 | −230.20 | 69.90 | 1.46 |
| G + U | 5 | 467.01 | 2.97 | 6.28 | −228.26 | 76.18 | 2.84 |
| A | 3 | 467.57 | 3.53 | 4.74 | −230.69 | 80.92 | 0.05 |
| A + O + U | 5 | 467.91 | 3.87 | 4.01 | −228.71 | 84.92 | 1.55 |
| G | 4 | 468.34 | 4.29 | 3.24 | −230.01 | 88.16 | 0.02 |
| A + O | 4 | 468.64 | 4.60 | 2.77 | −230.16 | 90.93 | 1.57 |
| G + A + U | 6 | 468.96 | 4.92 | 2.37 | −228.14 | 93.30 | 3.19 |
| G + O + U | 6 | 468.98 | 4.94 | 2.34 | −228.15 | 95.64 | 3.16 |
| G + O | 5 | 469.74 | 5.69 | 1.61 | −229.63 | 97.25 | 3.09 |
| G + A | 5 | 470.24 | 6.20 | 1.25 | −229.88 | 98.49 | 2.38 |
| G + A + O + U | 7 | 470.93 | 6.89 | 0.88 | −228.01 | 99.37 | 3.56 |
| G + A + O | 6 | 471.62 | 7.58 | 0.63 | −229.47 | 100.00 | 3.54 |

**Table 4** Detailed statistics of the three models within 2 units of AICc difference to the AICc-best model

| Independent variable | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| a.) U | | | | |
| (Intercept) | 0.848 | 0.274 | 3.099 | 0.002* |
| User experience (U) | 0.008 | 0.010 | 0.800 | 0.424 |
| b.) NULL | | | | |
| (Intercept) | 1.060 | 0.051 | 20.600 | < 0.001* |
| c.) O + U | | | | |
| (Intercept) | 0.747 | 0.321 | 2.325 | 0.020* |
| Openness to e-learning (O) | 0.021 | 0.035 | 0.605 | 0.545 |
| User experience (U) | 0.005 | 0.011 | 0.406 | 0.685 |

Table 4 shows detailed statistics for the three best models from the candidate set. We see that the variable user experience is part of both non-trivial winning models in the candidate set. Both models (U and O + U) have positive estimates, indicating an increase of 0.8% (0.5%) in the post-course test score for a 1-unit increase in the user experience variable (U) if all other variables are kept constant. However, the 95% confidence interval overlaps with zero as $p = 0.424$ ($p = 0.685$). Openness to e-learning (O) also had a positive effect with an increase of 2.1% in post-course test score for a 1-unit increase. The 95% confidence interval also contains zero ($p = 0.545$). The demographic variables gender (G) and age (A) are neither contained in any of the winning models, nor do the single predictor models with these two variables explain any notable amount of variability in learning success as measured by the post-course test scores as Nagelkerke's pseudo $R^2$ are practically zero for both models (i.e. 0.05% for A and 0.02 for G, respectively). Notably, the models scoring highest in terms of pseudo $R^2$ consistently score worst in terms of Akaike's information criterion (i.e. AICc).

## 4 Discussion

### 4.1 Descriptive and incidental findings

Even though we did not employ a strict pre-/post-design, in that we did not assess students' openness to e-learning before and after exposure, the difference in this variable between treatment and control group is striking. We find robust evidence that (1) sample means differed at a significance level of at least 0.1% as assessed by Welch's t-test and that (2) the evidence for the alternative hypothesis of unequal sample means is "decisive" as indicated by a Bayes factor larger than 100 (Kass & Raftery, 1995, 777). The difference in sample means with respect to openness to e-learning is such that those students that were randomly assigned to a Labster (i.e. e-learning) group appear to be decisively more open to e-learning than those in the control group. Because of the random assignment, we can exclude that this effect is due to self-selection bias and indeed a result of exposure to e-learning. It is also

unlikely that this finding might be a result of limited sample size. First and foremost, our sample is arguably not small when compared to other studies on university education in general and higher education in natural science, such as Hu-Au and Okita (2021) or Puntambekar et al. (2021), in particular. Second, post-hoc power for this study and this variable is 99.2% at the standard 5% significance level, which is well above the usual minimum goal of 80% power. Third, we complemented our analysis with the Bayes factor, which makes an actual statement about the relative evidence in favor of null and alternative hypothesis given the data, which is stronger than a mere evaluation of the probability of obtaining the data one has sampled if the null were true, i.e. the p-value. Finally, Bayesian methods are known to be robust tools even when analyzing smaller samples.

### 4.2 Learning success

Our results do not indicate any difference in learning success between treatment and control group as neither Welch's t-test nor the Bayes factor indicate any significant differences (cf. Figure 2), in agreement with earlier studies in comparable settings (e.g. Darrah et al., 2014; Hawkins & Phelps, 2013). We also do not find any differences when controlling for subject taught (cf. Figure 3), except for the subject of phosphorus cycle. This subject was the only subject in this study that did not have teaching materials in German in the Labster version. Students had to do the e-learning unit in English whereas conventional materials were available in German. Notably, this did not impact the test scores negatively when comparing them to the scores achieved in the other experiments, no matter whether they were conducted using traditional learning methods or using Labster. However, the conventional group doing the phosphorus cycle experiment performed significantly better than the Labster group, and significantly better than all other groups. One possible explanation for this is that for this particular subject the conventional teaching materials just happened to be more effective than their Labster counterparts. However, given the average first-semester student's level of English at Vechta University, we highlight that we actually found it surprising that we did not see any evidence of that particular subject being worse than average in terms of learning success. It remains unclear what impact the language of the virtual lab had on the performance of the treatment group. Comparable studies such as those by Bonde et al. (2014) and Makransky et al. (2016) do not address this aspect. One possible explanation for our result is that Labster is a multimedia product supported with images and actions and possible difficulties due to the language of instruction are not as significant because the units are less text-based. According to Leisen (2022), text-based representations are more abstract than pictorial or object-based representations (such as experiments). Labster uses both—pictorial elements and the possibility of experimenting—and thus also offers a lower degree of abstraction. Especially for students with low English-language aptitude, presentation levels of lower abstraction offer better chances of understanding content (Leisen, 2022). It seems possible that this effect occurred in the unit on the phosphorus cycle, so that any potentially negative effects from the use of English as language of instruction could have been counterbalanced.

However, as we did not find any general evidence that Labster improved learning success as compared to conventional teaching methods, some uncertainties as to the mechanisms at work remain, and point to areas of potential future research.

In summary, in terms of learning success, our study suggests two key take-home messages: First, Labster seems to perform about as good as conventional teaching methods for undergraduate biology education when it comes to learning success. A noteworthy caveat is that virtual labs cannot teach the practical lab skills that would be required in order to be able to work in an actual laboratory, so virtual labs can, for the currently foreseeable future, only ever complement university or higher education. Second, we did not find any evidence for the claim that e-learning would produce better learning outcomes than conventional teaching methods. However, this outcome does not contradict the findings of Makransky et al. (2016), Bonde et al. (2014) or Mosqueda (2023), who have demonstrated the advantages of Labster over other methods of learning. It shows that the learning success of virtual labs always depends on the situation as a whole (topic, context, etc.) and the object of comparison (e. g. conventional form of teaching or real lab). In our individual case, Labster offered itself as a further option for future course design. Its usefulness depends on various factors such as the the suitability of the units offered to our course topics. Especially when teaching abstract and difficult content, virtual labs have found to be an effective means of teaching (Byukusenge et al., 2022).

### 4.3 What makes (or breaks) a successful Labster unit?

The candidate modeling approach based on the information-theoretic paradigm found some limited evidence that a positive user experience (U) and a general openness to e-learning (O) mattered more than personal characteristics of students such as age (A) or gender (G) (Tables 3 and 4). However, this evidence is limited, because (1) we could not identify a clear winning model, i.e. several models are within two units of the AICc-best model and all but one model are within seven units, which still constitutes moderate evidence for the respective models (Burnham & Anderson, 2002); (2) the NULL model lies within two units of the AICc-best model, which is a strong indication that none of the candidate models are actually satisfactory models for the unknown data-generating process. The modest pseudo-$R^2$ values of all models in the candidate set, in particular of the three best models within two AICc units (U, NULL, O + U, cf. Table 3) are well in agreement with this conjecture. Nonetheless, the estimates that we find agree with intuition in that they generally indicate that a better user experience and more openness to e-learning both could have a positive effect on learning success as measured by the scores achieved in the 5-question post-course knowledge test. However, the lack of statistical significance indicates that we cannot reject the possibility to have found such estimates randomly (cf. Table 4). Taken together, the evidence that we present suggests that our a priori mental model of what determines learning success in e-learning settings seems to have been incomplete, which calls for further empirical and theoretical research.

Regarding e-learning in general, there have already been attempts to investigate success factors, for example by Abd El Halim and Elbadrawy (2021), who examined

key determinants of e-learning systems based on 17 hypotheses. They concluded that there is a positive correlation between system features, content quality, and perceived usefulness and system quality. However, the results of this study can only be applied to our investigations on Labster to a limited extent: This is because the term e-learning refers to a wide range of digital teaching methods. Within this pool, virtual labs in general and Labster in particular only represent a very small proportion. A more differentiated study of virtual laboratories in particular (in the context of biology teaching at universities) is therefore required in order to explain the determinants of success for our individual case. As Byukusenge et al. (2022) conclude, virtual labs are ultimately a tool whose effectiveness, as with any other tool, likely depends on how they are used, i.e. in which ways they are integrated into the classroom and how well they fit with course contents and the learning goals associated.

Lastly, in order to have a successful e-learning session, one first needs to have a session at all. This very simple principle is arguably not always, and not in all contexts, realizable. Consequently, many studies on e-learning and its potential for underprivileged and under-resourced areas of the world have pointed to inadequate internet access, especially in rural and underprivileged areas, as a substantial barrier to effective e-learning. For instance, Delita (2021) emphasizes that the lack of technology and reliable internet access prevents students from participating fully in online education. Similarly, Dutta and Smita (2020) report that interruptions in online classes due to poor internet speed and frequent power outages are common issues faced by students in Bangladesh. This sentiment is echoed by Kombe and Mtonga (2021), who identify lack of internet and electricity as critical barriers to e-learning for under-resourced students. Given the problems that surfaced on one occasion with establishing a reliable internet connection during the conduct of our study, internet reliability may also be an issue in underfinanced schools and universities in the Global North.

### 4.4 Strengths and limitations of this study

We introduced a bias in regression coefficients, because we were unable to account for the fact that up to four repeated measures on each individual were present in the data. Hence, our models did not account for within-subject correlation, because students did not have unique identifier codes that would have made it possible to identify them again in the next session. This might have biased the values of estimates and their errors, leading to increased type-I error rates (McCall & Appelbaum, 1973). On the other hand, because this type of bias holds for the entire set of candidate models, it is unlikely that the general conclusions from the regression analysis would change at all. In other words, the bias introduced might have affected statistical significances of estimates, but this will likely not alter the conclusion that all model candidates to some extent lacked in their approximation of the data-generating process, in particular regarding variability explained (cf. Table 3). The low pseudo-$R^2$ and the results of the candidate modeling process in general indicate that more research studying the success factors of e-learning in general and virtual labs in particular would help shedding light on this issue. On the other hand, the present

study quantitatively demonstrates the potential improvements in learning success given a still better understanding of the success factors of these technological elements of higher education.

## 5 Conclusions

This study extends the theoretical understanding of digital learning adoption by comparing virtual and traditional teaching methods in biology education. The results indicate that while learning outcomes remain consistent across both methods, virtual labs such as Labster significantly improve students' attitudes toward e-learning, which is critical to fostering engagement in digital learning environments. This supports virtual labs as a complementary tool, especially valuable when physical settings such as laboratories or traditional teaching are limited by cost or availability. Future research should examine content-specific adaptations and the influence of instructional language to ensure greater inclusivity and effectiveness in digital learning. By advancing practical knowledge about the use of virtual labs, this study provides further evidence for the development of targeted digital tools that improve accessibility and learning outcomes in higher education.

**Acknowledgements** A grant from University of Vechta's Central Study Commission is gratefully acknowledged. We thank Michael Ewig for helpful discussion of some aspects of the work presented here.

**Authors' contributions** All authors were involved in the preparation, implementation and evaluation of the study as well as in the writing of this article. All authors read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The study was financed by an internal grant from the University of Vechta.

**Data availability** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### Declarations

**Competing interests** The authors declare that they have no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References


Abd El Halim, H., & Elbadrawy, R. (2021). Examining drivers affecting E-learning system quality during COVID-19. *Total Quality Management & Business Excellence, 33*(11–12), 1422–1437.

Bhargava, P., Antonakakis, J., Cunningham, C., & Zehnder, A. T. (2006). Web-based virtual torsion laboratory. *Computer Applications in Engineering Education, 14*(1), 1–8.

Bonde, M., Makransky, G., Wandall, J., Larsen, M., Morsing, M., Jarmer, H., & Sommer, M. (2014). Improving biotech education through gamified laboratory simulations. *Nature Biotechnology, 32*, 694–697.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Byukusenge, C., Nsanganwimana, F., & Tarmo, A. P. (2022). Effectiveness of virtual laboratories in teaching and learning biology: A review of literature. *International Journal of Learning, Teaching and Educational Research, 21*(6), 1–17.

Chan, P., Van Gerven, T., Dubois, J.-L., & Bernaerts, K. (2021). Virtual chemical laboratories: A systematic literature review of research, technologies and instructional design. *Computers and Education Open, 2*, 1–17.

Chatterjee, S. (2021). Revolutionizing science education through virtual laboratories. In H. Biswas, S. Poddar, & A. Bhaumik (Eds.), *Advances in Science Education* (pp. 118–128).

Coleman, S. K., & Smith, C. L. (2019). Evaluating the benefits of virtual training for bioscience students. *Higher Education Pedagogies, 4*(1), 287–299.

Darrah, M., Humbert, R., Finstein, J., Simon, M., & Hopkins, J. (2014). Are virtual labs as effective as hands-on labs for undergraduate physics? A comparative study at two major universities. *Journal of Science Education and Technology, 23*, 803–814.

De Vries, L. E., & May, M. (2019). Virtual laboratory simulation in the education of laboratory technicians–motivation and study intensity. *Biochemistry and Molecular Biology Education, 47*(3), 257–262.

Delita, F. (2021). The challenges of E-learning implementation during the Covid-19 pandemic in senior high school. *Journal of Digital Learning and Education, 1*(3), 150–157.

Diwakar, S., Achuthan, K., Nedungadi, P., & Nair, B. (2012). Biotechnology virtual labs: Facilitating laboratory access anytime-anywhere for classroom education. In E. C. Agbo (Ed.), *Innovations in Biotechnology* (pp. 379–398).

Dutta, S., & Smita, M. (2020). The Impact of COVID-19 Pandemic on Tertiary Education in Bangladesh: Students' Perspectives. *Open Journal of Social Sciences, 8*, 53–68.

Glassey, J., & Magalhães, F. D. (2020). Virtual labs–love them or hate them, they are likely to be used more in the future. *Education for Chemical Engineers, 33*, 76–77.

Hawkins, I., & Phelps, A. J. (2013). Virtual laboratory vs. traditional laboratory: Which is more effective for teaching electrochemistry? *Chemistry Education Research and Practice, 14*(4), 516–523.

Heradio, R., De La Torre, L., Galan, D., Cabrerizo, F. J., Herrera-Viedma, E., & Dormido, S. (2016). Virtual and remote labs in education: A bibliometric analysis. *Computers & Education, 98*, 14–38.

Herzig, B. (2014). *Wie wirksam sind digitale Medien im Unterricht?* Bertelsmann Stiftung.

Hu-Au, E., & Okita, S. (2021). Exploring differences in student learning and behavior between real-life and virtual reality chemistry laboratories. *Journal of Science Education and Technology, 30*(6), 862–876.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Kolil, V. K., Muthupalani, S., & Achuthan, K. (2020). Virtual experimental platforms in chemistry laboratory education and its impact on experimental self-efficacy. *International Journal of Educational Technology in Higher Education, 17*(1), 1–22.

Kombe, C. L. M., & Mtonga, D. E. (2021). Challenges and interventions of eLearning for under resourced students amid covid-19 lockdown: A case of a Zambian public university. *Journal of Student Affairs in Africa, 9*(1), 23–39.

Leisen, J. (2022). *Sprachbildung und sprachsensibler Fachunterricht in den Naturwissenschaften*. Verlag W. Kohlhammer.

Makransky, G., Bonde, M. T., Wulff, J. S., Wandall, J., Hood, M., Creed, P. A., Bache, I., Silahtaroglu, A., & Nørremølle, A. (2016). Simulation-based virtual learning environment in medical genetics counseling: An example of bridging the gap between theory and practice in medical education. *BMC Medical Education, 16*(1), 1–9.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the Analysis of Repeated-Measures Designs: Some Alternative Approaches. *Child Development, 44*(3), 401–415.

Mosqueda, C. E. H. (2023). Effect of Utilizing Interactive Virtual Lab On Students Performance in Physics. *International Journal of Advanced Research, 11*(5), 1718–1741.

Puentedura, R. (2006). Transformation, technology, and education. Retrieved December 19, 2024, from http://hippasus.com/resources/tte/

Puntambekar, S., Gnesdilow, D., Dornfeld Tissenbaum, C., Narayanan, N. H., & Rebello, N. S. (2021). Supporting middle school students' science talk: A comparison of physical and virtual labs. *Journal of Research in Science Teaching, 58*(3), 392–419.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical. *Computing*. Retrieved December 19, 2024, from https://www.R-project.org/

Ray, S., & Srivastava, S. (2020). Virtualization of science education: A lesson from the COVID-19 pandemic. *Journal of Proteins and Proteomics, 11*(2), 77–80.

Schäfers, M. S., Schmiedebach, M., & Wegner, C. (2020). Virtuelle Labore im Biologieunterricht: Auswirkungen von Labster auf die Selbsteinschätzung von Schülerinnen und Schülern. *MedienPädagogik (Occasional Papers): Zeitschrift für Theorie und Praxis der Medienbildung*, pp. 140–167.

Thisgaard, M., & Makransky, G. (2017). Virtual Learning Simulations in High School: Effects on Cognitive and Non-Cognitive Outcomes and Implications on the Development of STEM Academic and Career Choice. *Frontiers in Psychology, 8*, 1–13.

Waldrop, M. (2013). Education online: The virtual lab. *Nature, 499*, 268–270.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.