Scaffolded Learning Assignments in University Machine Learning Education

A study into the effectiveness of assignment scaffolding

Maarten van der Weide[1]

Gosia Migut[1]

[1]**EEMCS, Delft Un**

Scaffolded
Learning
Assignments in
University
Machine Learning
Education

y, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Maarten van der Weide
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Mark Neerincx

Abstract

This study investigates the impact of scaffolded assignments on student learning, confidence, and the development of an empirical mindset in a Machine Learning (ML) course at TU Delft. Unlike traditional Computer Science subjects, ML requires an experimental approach, challenging students used to design-first methodologies. Through surveys of 25 students from the CSE2510 course, the study found that scaffolded assignments significantly enhance student confidence and perceived learning benefits, despite no positive correlation between the number of assignments completed and course grades. Qualitative feedback highlighted the value of scaffolded assignments in understanding the ML design process by providing structured guidance and enabling practice of specific sub-tasks. These findings suggest scaffolded learning is crucial in developing an experimental mindset and boosting student confidence in ML education. This study also proposes a second methodology for future research during an edition of the course to further explore this topic.

1 Introduction

Machine Learning (ML) has become a central part of the Computer Science discipline, leading universities to integrate more ML into their Computer Science Curricula. Traditionally, Computer Science courses focus on classical areas where a design-first approach is used. In these areas, problem-solving involves abstractly reasoning about the parameters, inputs and outputs to solve problems. However, ML is different from these classical parts of the curricula.

Designing ML models requires an experimental mindset. When designing these models, it is hard to reason about the concrete impact of design choices, such as the choice of hyperparameters. This necessitates a systematic experimental approach, where the design is empirically improved upon [1]. In contrast, when designing algorithms, a design-first approach is required, where the problem is abstractly solved by reasoning about the parameters, inputs, and outputs. This discrepancy is highlighted by Price, Salehi, Burkholder, *et al.* [2], who label such issues as ill-structured problems, stating that *they cannot be solved by deterministically following a set of instructions.* Although this definition does not exclusively pertain to the field of Machine Learning, it highlights the relation between problem-solving in novel situations and the gap between ML design and other more classical Computer Science courses. Because of this difference, the adoption of an empirical mindset is especially difficult for Computer Science students [3]. Consequently, it is imperative that students can sufficiently practice this explorative design approach.

One effective educational strategy to facilitate this practice is the use of scaffolded assignments. These assignments break down complex tasks into smaller, more manageable sub-tasks, providing intermediary hints or assistance at critical points. This method not only aids in developing an experimental mindset but also allows students to assess their understanding and build confidence in their problem-solving abilities Schiendorfer, Gajek, and Reif [3], Allen, McGough, and Devlin [4], and Otgon [5].

The CSE2510 course at TUDelft, an introductory course to Machine Learning taught in the second year of the bachelor Computer Science & Engineering, provides multiple optional scaffolded assignments[1]. These assignments are in the form of Jupyter notebooks that split up the bigger tasks into smaller pieces, which gives students the opportunity to reflect on parts of the ML design pipeline. The aim of this study is to investigate the relationship between these scaffolded assignments and student learning and confidence. By surveying students who have completed the course, the objective is to determine the effect of these assignments on student learning, both in terms of their summative course assessment and their perceived learning. Additionally, the students' confidence in their understanding of the course material and their confidence in applying the material knowledge outside of the course context is measured. Lastly, the perceived effectiveness of the assignments in developing a more empirical mindset is measured.

The methodology section describes in more detail the considerations and structure of the measurement instrument. The results section will contain an analysis and interpretation of the collected data. The discussion section will highlight the limitation and gaps that are present within the context of this study. The responsible research section will discuss and reflect on the ethical considerations relevant to this paper. Lastly, the conclusions section aims to sum up the aforementioned content and place it in the context of further work that can be done.

2 Related Work

Several studies have explored the use of scaffolded learning to enhance student outcomes in Computer Science education, both within and outside the context of Machine Learning.

Schiendorfer, Gajek, and Reif [3] propose scaffolded assignments to provide practice for students. These assignments are designed to help students gradually build the necessary skills to tackle more complex problems independently. The scaffolds serve as a temporary support system, which is gradually removed as students become more proficient. Their research suggests a teaching method that focuses on experimentation and understanding of ML concepts.

Price, Salehi, Burkholder, *et al.* [2] discuss the concept of ill-structured problems in education. They argue that these problems cannot be solved by following a set of deterministic instructions, thus requiring a more exploratory and iterative approach to problem-solving. Their work emphasizes the importance of developing an empirical mindset in students, which aligns closely with the experimental nature of ML tasks. By using scaffolded assignments, students are guided through the process of experimentation and iterative improvement, which is crucial for mastering ML techniques.

Furthermore, studies have demonstrated that scaffolded assignments not only help in developing an experimental mindset but also serve as a form of formative assessment [4], [5].

[1] Assignments can be accessed with TUDelft email *here*

This type of assessment provides students with immediate feedback on their performance, helping them identify areas of improvement and build confidence in their abilities. The research suggests that formative assessments, such as scaffolded assignments, can enhance student motivation and engagement, leading to better learning outcomes.

This study builds on these previous works by specifically examining the impact of scaffolded assignments in the context of the CSE2510 course at TU Delft. By surveying students, this research aims to provide insights into how scaffolded learning influences student outcomes and confidence in ML. The findings from this study are expected to contribute to the ongoing discussion about effective teaching strategies in ML education and the broader field of Computer Science.

3 Method

This section describes the structure and development of the method used to assess the relation between scaffolded assignments and the respective variables of interest; student learning, student confidence and the development of an empirical mindset. To this end, students who have participated in the CSE2510 course have been surveyed digitally[2]. Additionally, the design of the survey, with respect to the three domains, and the relevant statistical data is discussed. The choice of survey as measurement instrument was based upon the empirical nature of the research question. Respondent perception is inherent to their confidence, learning and mindset. Details on the ethical considerations and design choices made to this regard can be found in section 5.

3.1 Course background

The CSE2510 course is a second-year course taught at the Technische Universiteit Delft. This course is taught in the BSc. Computer Science & Engineering, is 10 weeks long and is worth 5 EC's (equaling a workload of 140 hours). It is the first Machine Learning course in the curriculum and thus serves as the first, formal, introduction to ML concepts and design methodologies for students. The course contents are described as follows: "The goal of the course is to acquaint students with the basic Machine Learning concepts and algorithms. Specifically, the course will cover parametric and non-parametric density estimation, linear and non-linear classification, unsupervised learning including clustering and dimensionality reduction, performance evaluation of predictive algorithms and ethical issues in machine learning."[3]

3.2 Participants

The study involved a group of 25 students, from the TU Delft, who have completed the CSE2510 course, in any year. Table 1 shows the distribution of the course edition respondents completed.

These students were recruited via personal networks. The survey was shared in student messaging groups of different academic years.

[2]Survey can be accessed with TUDelft email *here*, or be viewed in Appendix A

[3]www.studiegids.tudelft.nl, Year 23/24

Course edition	Number of respondents
2023/2024	6
2022/2023	9
2021/2022	7
Before 2021	3

Table 1: Year respondents completed the course.

Inclusion criteria

Given that the study focuses on the experience students have with the CSE2510 assignments, respondents who have not completed the course are excluded. Respondents are asked how many optional assignments they have completed during the course. In the case that a respondent had completed zero of these assignments, they continue the survey at question 15, they subsequently do not answer any question about the optional assignments.

Given the optional nature of all the survey questions, some respondents did not answer every question. Respondents who did not answer all questions are not excluded entirely. Respondents are only excluded for parts of the analysis where their missing answers are relevant.

3.3 Survey design

The contents of the survey can be broken up into the respective domains of the research question the questions aim to test. Additionally a part of the survey aims to empirically underline the findings of Schiendorfer, Gajek, and Reif [3], specifically, if students also perceive the proposed gap in mindset required by the Machine Learning course. Respondents were given access to a copy of the assignments available during the course, which they were free to use to refresh their memory.

Empirical verification

As a basis of this study lay the findings of Schiendorfer, Gajek, and Reif [3], which state that a gap in problem solving mindset is present in Computer Science students is present with respect to Machine Learning education. Students were asked if they agree with the statement that Machine learning model design requires a more empirical mindset, in comparison to other types of software development. This is an open question. The responses will be classified into three groups, agreeing, disagreeing and undecided. Afterwards, the given arguments of agreeing participants are classified . This serves to investigate the different reasons students might perceive the Machine Learning design process differently. The comparison of the given arguments to the work of Schiendorfer, Gajek, and Reif [3] serves as the empirical verification.

Respondents are asked to rate the usefulness of different available materials available during the course on a scale from 1 to 5. Work by Allen, McGough, and Devlin [4] suggests that students value practical exercises most out of available learning materials. The results are analyzed to investigate if that conclusion can be corroborated within this research group.

Learning

To assess the effect of scaffolded assignments on student learning the final course grade is used as a response variable. Furthermore, data is collected on the amount of optional lab assignments completed, if the participant completed the bonus assignment and if relevant their grade. On a scale of 1 to 5 students are asked to rate how often they complete optional materials in other courses. Lastly, the perceived impact of the assignments on their grade, on a scale from 1 to 5, is collected. A correlation between the amount of completed assignments and course grade is investigated.

The other collected variables are used to correct for intrinsic motivation if necessary. Assuming students exist that partake in all optional assignments, have a consistently high GPA, independent of provided material quality. These variables can then be useful to see the measure of correlation without this group.

Confidence

In their work, Niemi, Nevgi, and Virtanen [6] show that Web Based Pedagogical Tools (WBPT) can serve as effective scaffolds, specifically content creation and delivery tools, in the case of this study, Jupyter Notebooks, are reported as supporting self-evaluation, task strategies and goal setting. Self-evaluation, or formative assessment, is shown to be beneficial to student learning [4], [5]. To this end students were asked to self-report to what extend they feel that the assignments provided them with an insight into their understanding of course material on a scale from 1-5.

Additionally, a part of the Instructional Materials Motivation Survey (IMMS) was used. The IMMS is a scale designed by Keller [7]. It is a validated scale that consists of four sub-scales that can be tested independently [8]. The sub-scales respectively test the constructs Attention, Relevance, Confidence, and Satisfaction. In the context of this study only the confidence sub-scale was included. The selected sub-scale consists of 9 questions, these correspond with question 11.1 through 11.8 and 11.12 of the survey. The scale does not have a norm, meaning that the results can not be classified as *low* or *high* [7]. To interpret the results the responses will be used to calculate the Cronbach's α to assess the internal consistency of the results. This serves as a way to verify the responses are measuring a related construct, namely confidence.

The IMMS results and the perceived insight gained into course understanding mentioned earlier will be used to determine if a correlation exists between these two variables. Namely, if students report that the assignment was perceived as formative, it is expected that this also increases student confidence[4], [5]. To this end the Cronbach's α of the average IMMS response, together with the perceived gained insight will be calculated. The result will be used as a measure of relation between the internal constancy between the IMMS survey and perceived gained insight into material understanding. This is possible additional indicator of the effectiveness of the IMMS scale in this context, which further research can build upon.

Empirical mindset

As stated in the works of Schiendorfer, Gajek, and Reif [3] and Hazzan and Mike [1], (scaffolded) practicing is imperative in supporting development of the required design mindset. To test this respondents were asked multiple open-ended questions. As stated earlier, they were asked about their opinion on the existence of the gap in mindset found by Schiendorfer, Gajek, and Reif [3].

Subsequently, respondents are asked the following question: *Did the optional lab assignments help you better understand the design process involved in creating (some) machine learning models? If so, in what ways?* (question 13). This question aims to analyze perceived benefit with respect to the development of the intended mindset. Agreeing responses to this question will be classified based on their arguments. This serves as a way to verify that (scaffolded) assignments are experienced as having a positive influence on design process understanding. The classifications are used to investigate to what extent the arguments coincide with empirical parts of the ML design process.

Lastly the respondents will be asked: *Do you believe that the step-by-step structure of the optional lab assignments enhanced your learning experience? If so, please describe how.* Similar to the previous question, this has the aim to identify what students appreciate about the scaffolding specifically. Responses are also contextualized with respect to the experimental nature of ML design present in the assignment to investigate the relation between scaffolding and the perceived development of the required empirical mindset.

4 Experiment during course edition

This study primarily relies on retrospective self-reporting by students who have already completed the CSE2510 course. While this approach provides valuable insights into the students' perceptions and the potential impact of scaffolded assignments on their learning, confidence, and development of an empirical mindset, it has inherent limitations. This consideration is discussed in further detail in section 7. This section proposes an alternative methodology that can be used during future editions of the course, with the aim of overcoming this limitation and gaining a more comprehensive understanding of the effects of scaffolded learning assignments.

4.1 Participants

Participants will include students enrolled in the CSE2510 course. Under the assumption that the research is conducted or supervised by responsible course staff, subsections of relevant data can be collected from all enrolled students. This data includes course results, as a grade, and insight into assignment completion. Other relevant data can be collected from voluntary participants, this will be in the form of surveys.

4.2 Learning

This described method will investigate the effect of scaffolded assignments on learning, as measured by summative course results, perceived learning, and various formative assessment techniques.

By collection enrolled students' summative course result, exam submission and per-assignment engagement, the impact of scaffolded assignments on student learning is investigated. Firstly, student assignment engagement can be correlated with respective course result. Given that this data can be collected and anonymously aggregated from all enrolled students, a significant response size is available. Additionally, the optional nature of the assignments should provide for a significant group of students that have not or have only partially engaged in the provided assignments. This creates multiple classes of students with each different amount of assignment completion. This data can then be used to more accurately measure the relation between multiple classes of assignment completion and summative course results.

By collecting data on the results of each exam question, the per-topic effect of the assignments can be investigated. It is important to note that scaffolded assignments should align closely with the intended learning outcomes (ILOs) and assessments (constructive alignment). To achieve this, the individual exam questions can be related to the topics covered in the assignments. This mapping, along with the level of engagement with the assignments, is utilized to more precisely measure the impact of scaffolded assignments on summative results on a per-question basis.

If possible, student GPA and optional assignment completion in other courses can be collected as a correcting variable that signifies intrinsic motivation. Students who have a high GPA and often complete optional assignments can be expected to receive a high summative result. The effect of the assignment could be small for this group. In case this data cannot be collected an optional survey can be distributed amongst the students after the course. In this survey students self-report their GPA, CSE2510 grade and engagement in optional materials in other courses.

4.3 Confidence

Future research during an edition of the course enables the collection of data at intervals between optional assignments. The Instructional Materials Motivation Survey (IMMS) scale utilized in this study can again be administered. The IMMS results presented in this paper can be used as a comparison measure in all future research into CSE2510, or related materials research.

The IMMS can be administered after each individual assignment. This approach allows for the measurement of confidence levels specific to each assignment, facilitating comparisons between assignments and results presented in this paper. Additionally, administering this survey after each assignment could give a finer result, given that respondents do not have to reflect on an assignment that they have engaged in long ago. The other domains of the IMMS scale, namely attention, satisfaction and relevance, can be included in this measurement. This is not directly related to the research question in this paper, but can provide valuable data for the course staff to possibly further improve the quality of the course.

4.4 Empirical Mindset

To qualitatively measure the problem solving skills of students Price, Salehi, Burkholder, *et al.* [2] developed the *Decision-Based Assessment Template*. This template distinguishes itself by viewing a 'skilled problem solver' as someone who can not only produce the correct solution to a problem, but also have an 'adaptive expertise' such that they can solve problems in a range of novel situations. This is in line with the definition used in this paper of an 'empirical mindset'.

The template outlines the following process.

1. Provide an authentic problem scenario
2. Ask a series of questions that require test-takers to make decisions about problem definition and planning how to solve
3. Provide more information, which includes both information an expert would seek and irrelevant information novices might encounter or seek, and/or more specific criteria needed to define the problem to be solved;
4. Ask a series of questions that require decisions about interpreting information and drawing conclusions
5. Have test-takers choose and reflect on their solution.

It is important to note the overlap between this assessment process and scaffolded assignments. The core idea is similar but differs in the rigorous structure CSE2510 assignments have and also asking the student to reflect on the assignment as a whole.

Utilizing this template the empirical mindset of students will be assessed post course as follows. An assignment similar to the existing ones available will be provided. This assignment however more loosely constraints the student. This implies making the scaffolds less constraining and letting the student make more impact design decisions. The assignment will still have a serial nature trough the scaffolds and will provide "both information an expert would seek and irrelevant information novices might encounter" between the scaffolds. Progressing through this assignment, more constraints can be added (e.g. performance, ethical considerations). Finally, students will be asked to reflect on their process and solution.

This paper presents an example assignment following this template[4]. This assessement includes all the mentioned steps but excludes a scoring rubric, which, in accordance to the template, needs to be developed based on experts engagement with the assessment.

This process gives an insight into the decisions students make in novel situations, subsequently assessing their empirical mindset or 'adaptive expertise'. This test will be a post-test. By selecting a group of respondents that have engaged in the scaffolded assignments to different degrees, this assessment is used to investigate the effect of the scaffolded assignments on the development of an empirical mindset.

5 Responsible Research

This section serves as a reflection on the ethical aspects of the research and the reproducibility of the presented methods.

5.1 Ethical Considerations

The following measures were taken to ensure the ethical integrity of this study and to adhere to the human research ethics guidelines presented by the TU Delft[5].

[4]Proposed assessment can be accessed with TUDelft email here
[5]Guidelines can be found here

Reason	Count
Per-problem approach[1]	6
Parameter tweaking	2
Performance evaluation	2
Experiment to solve new problems	1
None	2

Table 2: Classification of agreeing arguments
[1] Includes both experimenting with models and datasets.

- **Informed consent**: Participants of this survey were provided with clear information about the purpose of the research and their rights as participants. Informed consent was obtained from all participants before any data was collected. The informed consent statement can be found in Appendix A.
- **Anonymity**: Participants provided only data relevant to the research. The collected data is only presented as aggregate, to minimize possible risk of re-identification of respondents.
- **Confidentiality**: The relevant data was collected via the TU Delft Microsoft forms environment. Non TU Delft accounts were neither allowed as respondents or used in the creation and management of the survey.
- **HREC and Data Management Plan**: In accordance to TU Delft policy the above points, and more, were aggregated into a HREC form and submitted to the TU Delft for approval.

5.2 Reproducibility

Ensuring reproducibility is essential for the transparency and integrity of this research. For this reason the methods used in this paper are carefully documented. This includes data collection methods, data analysis methods and relevant information about the participants.

As mentioned, to minimize the risk of re-identification of participants, the collected dataset is not published. The collected data is however presented in an aggregated form. Future research can refer to the data in this aggregated form and thus this does not limit the reproducibility.

6 Results

In this section the results, obtained by collecting and analyzing the data as described in section 3, are presented and discussed.

Empirical verification

To verify the findings by Schiendorfer, Gajek, and Reif [3], that a gap in problem solving mindset is present, the respondents were asked to provide their opinion on the difference in required mindset for Machine Learning Engineering versus other types of Software Engineering. 8 respondents have left this question blank. From the remaining 17 respondents, 12 agreed with the statement, 4 disagreed and 1 agreed only in the specific case of deep learning.

Table 2 shows a classification of the given arguments by agreeing respondents. Performing a z-test with $p = 0.05$ yields that significantly more respondents agreed with the statement. The found argument classifications also coincide with the work of Schiendorfer, Gajek, and Reif [3], who picked the following 3 important skills:

- Data-splitting and correct usage of data subsets, given that this determines model validity
- Tuning of hyperparameters
- Understanding numerical optimizers.

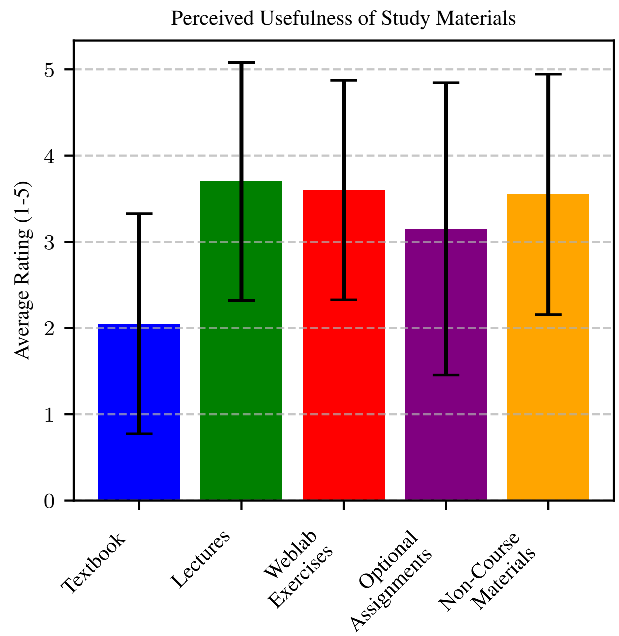

Figure 1: Reported usefulness of available study materials, on a scale of 1-5. The error bars represent the standard deviation.

The participants most valued study materials are shown in Figure 1. Findings by Allen, McGough, and Devlin [4] suggest that students value practical exercises in Machine Learning courses most. In the presented results the same conclusion cannot be drawn. The respondents rated lectures, Weblab exercises and Non-course materials higher than the lab assignments.

Learning

The results pertaining to investigating learning benefit are shown aggregated in Table 3. The count column illustrates the subsection of respondents that answered the respective question. Responses missing an answer for a particular variable are not counted. Only the first exam attempt result is counted.

Based on the gathered data, there is a correlation coefficient of -0.01 between the course grade and the number of assignments completed. This result suggests that, based on the sample, there exists almost no correlation between these two variables. When excluding data points with zero completed assignments, the coefficient becomes -0.13, which implies that there exists a small negative correlation.

Variables	Count	Mean (SD)
Exam grade	25	7.1 (1.65)
Assignments completed[1]	23	3.67 (2.85)
Perceived impact on grade[2]	16	3.56 (1.46)
Bonus Grade[3]	5	10 (0)
Attitude towards optional material[2]	23	3.17 (1.07)

Table 3: Findings statistics

[1] Out of a total of 7.
[2] All respondents received full marks.
[3] On a scale of 1-5

The absence of a positive correlation is in contrast to similar research results. Simons and Klein [9] find that scaffolding improves student learning, as measured by grade, in comparison to the absence of scaffolding. Additionally, the formative nature optional assignments have, even without scaffolding, should suggest a positive impact on course grade [4], [5].

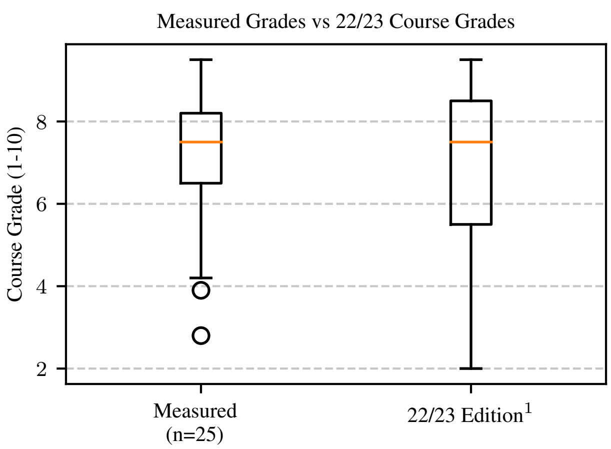

Figure 2: Boxplot of measure grades and grades from the 22/23 course edition.

[1] Distribution as published to students who participated in the 22/23 edition via Osiris.

The discrepancy between previous work and the presented findings could be attributed to an unrepresentative dataset. This is further supported by the observation that the dataset only contains 4 data points of failing grades. Figure 2 illustrates the centering of the measured grades, in comparison to results from a previous course edition. High cohesion of respondent's grade, in an already small sample size, makes it hard to identify possible correlations. Including only the subset with a passing grade, the correlation between grade and number of assignments done becomes 0.1.

Confidence

The data collected using the IMMS scale is shown in Table 4. In total 13 out of 25 respondents fully answered these questions, responses not answering every IMMS question are excluded. The question numbers correspond to the ordering described in section 3. These results correspond to a Cronbach's α of 0.78. This implies a similar internal consistency to the results obtained by Keller [7] ($\alpha = 0.81$), and the results obtained by Loorbach, Peters, Karreman, *et al.* [8] in their validation ($\alpha = 0.73$). Respondents were also asked to what extend they think the assignments provided them with insight into their own understanding of the course material on a scale from 1-5 ($\mu = 3.54$ $\sigma = 1.45$). Correlating the average respondents IMMS result with this variable gives a correlation coefficient of 0.67 ($p = 0.01$). This finding shows that formative assessment provides students with confidence, but at the same time contradicts the findings presented about learning, as this increased confidence should lead to higher performance[4], [5].

Questions	Mean	Standard deviation
1	2.92	1.04
2	3.85	1.07
3	3.46	1.20
4	3.54	1.05
5	3.92	0.95
6	3.38	1.33
7	3.46	1.13
8	3.15	1.28
9	1.92	1.12

Table 4: IMMS survey results (n=13)

Reason	Count
Experiencing the entire design process	6
Experimenting with different approaches	2
Experiencing evaluation and improvement	1
Providing context to theory	3

Table 5: Classification of agreeing arguments

Furthermore, correlating students perceived impact of the assignments on their grade with the average IMMS result gives a correlation coefficient of 0.65 ($p = 0.01$). Adding both these variables yields a Cronbach's α of 0.84, which implies that there exist some consistency between the measurement of confidence, perceived effect of materials on grade and perceived effect of assignments on grade.

Empirical mindset

Respondents were asked their opinion to the following question: *Did the optional lab assignments help you better understand the design process involved in creating (some) machine learning models? If so, in what ways?* To this question 17 respondents gave an answer. Out of 17 total responses, 11 agreed, 5 disagreed and 1 was undecided. The given arguments are shown in Table 5. The given ways in which respondents perceived benefit form the assignments coincides with largely with the practice needed to develop the required empirical mindset. Here practicing the entire design process is mentioned most, which aligns with findings from Hazzan and Mike [10], who highlights the importance of practicing in a problem based environment, to experience the entire design process.

Subsequently students were asked the question: *Do you believe that the step-by-step structure of the optional lab assignments enhanced your learning experience? If so, please* *describe how.* Out of 17 responses to this question, 13 agreed, 3 were undecided and 1 disagreed. All the agreeing respondents were unified in their argumentation. Namely, *Scaffolding provides guidance*. Respondents reported that the structure made it easier to focus on bits of the design process, which was perceived as less overwhelming. Additionally, the scaffolding made it possible to test and evaluate individual parts of the code, thus providing explicit practice with these sub-tasks. Three respondents explicitly mentioned that this structure gave them a better picture of how to create Machine Learning models and that it better helped them understand the way in which the individual concepts work together.

7 Discussion

As previously touched upon, results presented both align with, and diverge from, relevant publications. To this end, this section serves to aggregate and further contextualize important points of consideration pertaining to the results presented.

Part of the method aimed to verify empirically findings that serve as a basis for this research. Results aligned with previous findings regarding the need for an empirical mindset in Machine Learning. However, contrary to expectations[4], [5], [9], the collected data showed no direct correlation between the completion of scaffolded assignments and course grades. This does not conclusively contradict relevant findings and could be attributed to other factors, namely, the relatively small sample size and possibly unrepresentative sample.

Both these factors can be substantiated by the collected data, namely due to the low variance in grades of respondents, which can be a sign of an unrepresentative dataset. Given a larger more representative dataset findings could very well align with relevant findings.

A positive relation between scaffolded assignments and student confidence was established. Students reported gaining valuable insight from these assignments, which contributed to their confidence. Additionally, qualitative responses highlighted the benefit of scaffolded assignments in providing hands-on experience with the design process.

This study primarily relies on the recollection of respondents, as this study was not preformed during a CSE2510 course edition. This could have an impact on the results and could not be attributed for within the confines of this research. To this end, this paper presents an additional methodology that could be executed during future editions of the course to validate and extend upon the findings of this paper. This method, combined with larger and more diverse samples can greatly increase and contextualize knowledge of the topic at hand.

8 Conclusions

This research explored the impact of scaffolded assignments on student learning, confidence, and the development of an empirical mindset in the Machine Learning course CSE2510 at TU Delft. The primary research questions were:

- Do scaffolded assignments enhance student learning outcomes in Machine Learning?
- Do these assignments improve student confidence in their understanding of the course material?
- Do scaffolded assignments help in developing the required empirical mindset for Machine Learning?

The findings indicate that while scaffolded assignments do not have a significant positive correlation with improved course grades, they do substantially enhance student confidence and their understanding of the design process in Machine Learning. The majority of students agreed that the ML design requires a different mindset compared to other Computer Science subjects, emphasizing the need for a per-problem approach and the importance of parameter tweaking and performance evaluation. The scaffolded assignments provided structured guidance, making complex tasks more manageable and enabling students to practice specific sub-tasks effectively.

This study contributes to the understanding of how scaffolded learning can support the development of an empirical mindset in ML education. It illustrates the importance of further research into the educational side of ML and to this end provides a detailed methodology to further research the topic.

Further research could also examine the specific components of scaffolded assignments that are most effective, allowing educators to refine these tools to maximize their impact on both learning outcomes and student confidence.

References

[1] O. Hazzan and K. Mike, "The pedagogical challenge of machine learning education," in *Guide to Teaching Data Science: An Interdisciplinary Approach*. Cham: Springer International Publishing, 2023, pp. 199–208, ISBN: 978-3-031-24758-3. DOI: 10.1007/978-3-031-24758-3_13. [Online]. Available: https://doi.org/10.1007/978-3-031-24758-3_13.

[2] A. Price, S. Salehi, E. Burkholder, *et al.*, "An accurate and practical method for assessing science and engineering problem-solving expertise," *International Journal of Science Education*, vol. 44, no. 13, pp. 2061–2084, 2022. DOI: 10.1080/09500693.2022.2111668.

[3] A. Schiendorfer, C. Gajek, and W. Reif, "Turning software engineers into machine learning engineers," in *Proceedings of the First Teaching Machine Learning and Artificial Intelligence Workshop*, B. Bischl, O. Guhr, H. Seibold, and P. Steinbach, Eds., ser. Proceedings of Machine Learning Research, vol. 141, PMLR, 14 Sep 2021, pp. 36–41. [Online]. Available: https://proceedings.mlr.press/v141/schiendorfer21a.html.

[4] B. Allen, A. S. McGough, and M. Devlin, "Toward a framework for teaching artificial intelligence to a higher education audience," *ACM Transactions on Computing Education*, vol. 22, no. 2, pp. 1–29, 2021. DOI: 10.1145/3485062.

[5] J. Otgon, "The study of the relationship between students' self-assessment and their attitudes towards learning," in *Quality Assurance in Higher Education International Conference (QAHE 2022)*, Atlantis Press, 2022, pp. 121–129.

[6] H. Niemi, A. Nevgi, and P. Virtanen, "Towards self-regulation in web-based learning," *Journal of Educational Media*, vol. 28, no. 1, pp. 49–71, 2003. DOI: 10.1080/1358165032000156437.

[7] J. M. Keller, *Motivational Design for Learning and Performance*. Springer US, 2010. DOI: 10.1007/978-1-4419-1250-3.

[8] N. Loorbach, O. Peters, J. Karreman, and M. Steehouder, "Validation of the instructional materials motivation survey (imms) in a self-directed instructional setting aimed at working with technology," *British Journal of Educational Technology*, vol. 46, no. 1, pp. 204–218, 2015. DOI: https://doi.org/10.1111/bjet.12138. eprint: https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12138. [Online]. Available: https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12138.

[9] K. D. Simons and J. D. Klein, "The impact of scaffolding and student achievement levels in a problem-based learning environment," *Instructional Science*, vol. 35, no. 1, pp. 41–72, 2006. DOI: 10.1007/s11251-006-9002-5.

[10] O. Hazzan and K. Mike, "Teaching methods for machine learning," in *Guide to Teaching Data Science: An Interdisciplinary Approach*. Cham: Springer International Publishing, 2023, pp. 235–249, ISBN: 978-3-031-24758-3. DOI: 10.1007/978-3-031-24758-3_16. [Online]. Available: https://doi.org/10.1007/978-3-031-24758-3_16.

Appendix A

Survey administered to participants. Questions 6, 7, 8 and 9 are part of a different research project.

Survey on study methods and assessment in Machine Learning

You are being invited to participate in two research studies titled "The influence of assessment methods on students' performance in an exam on k-means clustering" and "Scaffolded learning assignments in university machine learning education". These studies are being done by Madeline El Aissati and Maarten van der Weide respectively, for the bachelor thesis in CSE.

The purpose of the first research study is to identify the effect of the assessment method on the performance of a student, specifically related to the Machine Learning topic "k-means clustering".

The purpose of the second research study is to measure the effect of the optional assignments provided in the bachelor machine learning course (CSE2510) on student performance, confidence and mindset towards ML development.

The survey will provide us with background data on students' study methods, their grades and their opinion on the Machine Learning assignments. Specifically we are looking for **students who took the ML course.** You do not need to have a passing grade: we are interested in the process of taking the course!

The survey will take you about 15 minutes to complete and will be conducted between 6/05/24 - 03/06/24.
The data will be used for a bachelor thesis in the field of Machine Learning education, and the de-identified results of the study will be published on GitHub.

As with any online activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by storing your personal data (gender and age) using TU Delft OneDrive, it will not be shared beyond the study team. Your gender and age will be recorded but will not be mentioned individually in the findings of the study: merely a summary of the gender division and the mean age will be reported. Your grade for the course will **not** be tied to your personal details; you cannot be identified by this grade. No other personal data will be collected. Any identifiable personal data you provide will be destroyed after 3 months.

Your participation in this study is **entirely voluntary and you can withdraw at any**

time, without giving a reason. You are free to omit any questions. As your survey results are anonymous no specific data can be deleted.

Your survey answers may be quoted anonymously in the results of the study.

By filling out the survey you agree to the described use of your personal

This section focuses on your study methods and obtained grade for the CSE2510 Machine Learning course.

1

I agree with the described use of my personal data and understand I can withdraw from the study at any given time.

- ◯ Yes
- ◯ No

2

Gender

- ◯ Male
- ◯ Female
- ◯ Non-binary
- ◯ Prefer not to say

3

What is your age?

- ◯ 18-25
- ◯ 26-30
- ◯ > 30

4

I took the Machine Learning course from year 2 of the Bachelor CSE in:

- ☐ 2023/2024
- ☐ 2022/2023
- ☐ 2021/2022
- ☐ Before 2021
- ☐ I did not take this ML course (you can exit the survey in this case)

5

What was your final grade for the course?

If you took the course multiple times, please report all final grades. For example: final exam 1: 4.5, final exam 2 (resit): 5.9

6

Please describe below the study methods you used throughout the course (examples: summarizing, reading the book, watching videos, flash cards etc.). Be **as specific** as possible!

7

If you took the final exam more than once, did you change your study methods at all? If yes, please elaborate on how you studied for each exam.

For example: final exam 1: summarized the book, resit exam: used the assignments etc.

8

To what extent do you feel the final exam reflected the learning objectives of the course?

If you took the resit, you may choose one of the two exams (**please specify** in the next question which exam you chose)

Study Goals	After succesfully completing this course, the student is able to: - explain the basic concepts and algorithms of machine learning and their un - implement, apply and evaluate basic ML algorithms in Python. - explain the concept of and identify (implicit) bias in data and ML algorithms

1	2	3	4	5	6	7	8	9	10

Not at all Completely

9

Which exam did you base your previous answer on? Please provide more information on your rating.

For example: my rating is based on the resit of ML 2022/2023. I rated a 5 on the previous question because ...

The following section of the survey will focus on your experience with the optional lab assignments provided during the CSE2510 (Machine Learning) course. These lab assignments were Python notebooks designed to reinforce the taught material and provide practical exercises. Throughout the rest of survey, when referring to "optional lab assignments," we are specifically talking about these Python notebooks.

Feel free to look at the optional lab assignments again as a refresher: https://tud365-my.sharepoint.com/:f:/g/personal/mvvanderweide_tudelft_nl/EiKpFS2pmcNIg4-gQsfJidUBun3fcxIkolwhnxzfqPYW2A?e=wi4qwo

10

During the course, there were 7 topics with optional lab assignments. How many of these did you complete? In case you completed none, you can continue to question 14.

It might be helpful to look at one of the assignments again as a refresher (https://tud365-my.sharepoint.com/:f:/g/personal/mvvanderweide_tudelft_nl/EiKpFS2pmcNIg4-gQsfJidUBun3fcxIkolwhnxzfqPYW2A?e=wi4qwo)

Number must be between 0 ~ 7

11

To what extent do you agree with the following statements about the optional assignments provided in the Machine Learning course?

It might be helpful to look at one of the assignments again as a refresher (e.g. https://tud365-my.sharepoint.com/:f:/g/personal/mvvanderweide_tudelft_nl/EiKpFS2pmcNIg4-gQsfJidUBun3fcxIkolwhnxzfqPYW2A?e=wi4qwo)

	Not true	Slightly true	Moderately true
When I first looked at the optional assignments, I had the impression that they would be easy for me.	○	○	○
The material in the optional assignments was more difficult to understand than I would like for it to be.	○	○	○
The introductory information in the optional assignments made it clear what I was supposed to learn.	○	○	○
As I worked on the optional assignments, I was confident that I could learn the content.	○	○	○
The exercises in the optional assignments were too difficult.	○	○	○

After working on the optional assignments, I was confident that I would be able to pass a test on it.	○	○	○
I could not really understand quite a bit fo the material in the optional assignments.	○	○	○
The good organization of the content helped me be confident that I would learn the material.	○	○	○
The optional assignments positively impacted my course grade.	○	○	○
The optional assignments provided me with an insight on my into my understanding of the course material.	○	○	○
The optional assignments gave me more confidence about my ability to apply machine learning theory outside of the course.	○	○	○
The optional assignment contained so much information, that it was had to pick out and remember the	○	○	○

important
points

12

In your opinion, does designing machine learning models require a more
experimental approach compared to other types of software development
(e.g., algorithm design, traditional software design)? Please explain.

13

Did the optional lab assignments help you better understand the design process involved in creating (some) machine learning models? If so, in what ways?

It might be helpful to look at one of the assignments again as a refresher (e.g. https://tud365-my.sharepoint.com/:f:/g/personal/mvvanderweide_tudelft_nl/EiKpFS2pmcNIg4-gQsfJidUBun3fcxIkolwhnxzfqPYW2A?e=wi4qwo)

14

Do you believe that the step-by-step structure of the optional lab assignments enhanced your learning experience? If so, please describe
how.

It might be helpful to look at one of the assignments again as a refresher (e.g. https://tud365-my.sharepoint.com/:f:/g/personal/mvvanderweide_tudelft_nl/EiKpFS2pmcNIg4-gQsfJidUBun3fcxIkolwhnxzfqPYW2A?e=wi4qwo)

15

On a scale from 1-5 how useful did you find the following resources during the course? (5= very useful)

	1	2	3
Textbook	○	○	○
Lectures	○	○	○
Weblab exercises	○	○	○
Optional lab assignments	○	○	○
Non-course materials	○	○	○

16

Did you complete, and receive a grade, for the bonus assignment provided during the course?

○ Yes

○ No

17

What grade did you receive for the bonus assignment?

The value must be a number

18

In your experience with other courses, how often do you typically complete optional assignments?

Optional assignments refer to any non-mandatory, available course work. (e.g. knowledge quizzes, bonus assignments)

- Always
- Often
- Sometimes
- Rarely
- Never

19

Do you have any additional opinions, tips or feedback you want to share about the course?