Author(s): Lahme, Simon Z.; Klein, Pascal; Lehtinen, Antti; Müller, Andreas; Pirinen, Pekka; Rončević, Lucija; Sušac, Ana

Title: Evaluating digital experimental tasks for physics laboratory courses

Year: 2023

Version: Published version

Evaluating digital experimental tasks for physics laboratory courses

Simon Z. Lahme[1,a], Pascal Klein[1], Antti Lehtinen[2,3], Andreas Müller[4], Pekka Pirinen[2], Lucija Rončević[5], Ana Sušac[5]

[1] Faculty of Physics, Physics Education Research, University of Göttingen, Friedrich-Hund-Platz 1, 37077, Göttingen, Germany
[2] Department of Physics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland
[3] Department of Teacher Education, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland
[4] Faculty of Sciences, Department of Physics, University of Geneva, Boulevard du Pont d'Arve 40, 1211, Genève, Switzerland
[5] Department of Applied Physics, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000, Zagreb, Croatia
[a] simon.lahme@uni-goettingen.de

Abstract

As physics laboratory courses are an integral part of studying physics, many approaches have been pursued to evaluate their quality e.g., regarding the improvement of conceptual understanding, the students' motivation, or the acquisition of adequate concepts about experimental physics. So far, most approaches either evaluate laboratory courses in their entirety like a course evaluation or focus on the students' development of (specific) competencies. However, even though experimental tasks are the backbone of any laboratory course concept, specific instruments to evaluate individual experimental tasks are missing. Both approaches mentioned above are unsuitable for that aim since typical laboratory courses consist of multiple tasks and the development of competencies takes place on a larger time scale than the execution of individual tasks. Thus, as part of the Erasmus+ project DigiPhysLab (Developing Digital Physics Laboratory Work for Distance Learning), we developed a questionnaire to explicitly evaluate the quality of an individual experimental task. The questionnaire has been discursively developed and softly validated within our project group and is now available in four languages. In this contribution, we share our ideas behind and our experiences with the use of this instrument for piloting experimental tasks.

1. Motivation

University physics laboratory courses usually follow the approach of task-based learning (cf. Müller & Brown, 2022); this means that experimental tasks are understood as "an idea/concept of an experiment- and task-based learning environment with materials like task instruction sheets, lab equipment, etc." (Lahme et al., 2023b, p. 2) are the main learning opportunity for students in this course format. In the sense of The Offer-and-Use-Model (Helmke, 2007) instructors develop and prepare suitable experimental tasks. This offer can for example consist of written task instructions, prepared experimental equipment, or guidance during the execution of the experimental tasks. The students use the experimental tasks (often in pairs or small groups) for their own learning process. They conduct the experiments with the given equipment following the provided task instructions and by this, (hopefully) the desired outcome is achieved (e.g., reinforcing a concept, improving experimental skills, etc.). The quality and especially the outcome of this learning process depend on the quality of the experimental tasks so on the quality of the instructors' offer for their students. Thus, instructors need to be careful when designing new experimental tasks for their physics laboratory courses and after a first implementation, they should evaluate and respectively re-design their experimental tasks.

As described by Lahme et al. (2023b), instructors should follow six design principles during this design process: They need to think of their target group and their learning objectives, need to come up with a task conception and explicit materials needed for the conduction of the experimental task and implement the tasks in their lab courses while acknowledging existing circumstances. Based on these six design principles an iterative process of design and re-design in the sense of an Action Research approach (Costello, 2003) can follow to continuously improve the experimental task. So, after each implementation instructors should evaluate and improve their task conception and designed materials based on students' feedback (and their observations). A systematic evaluation especially with larger groups of students can best be done with a questionnaire. To our knowledge, no questionnaire exists that primarily addresses the evaluation of individual experimental tasks for physics laboratory courses. In this contribution, we present our draft of such a questionnaire developed in the Erasmus+ project DigiPhysLab (see Lahme et al., 2022).

2. Need for a new questionnaire

As educational research on physics laboratories has been conducted for several decades, some instruments have already been developed to evaluate physics lab courses. They belong to two general approaches that have already been pursued. First, instruments have been developed to evaluate a lab course in its entirety i.e., with no specific focus on an individual task but with a general evaluation of the whole course program e.g., at the end of a semester. An example of such an instrument is the PraQ-questionnaire (Praktikumsqualitätsfragebogen; Rehfeldt, 2017). It consists of items to be rated by the students at the end of a whole lab course that addresses the students' perceived learning gain regarding the acquisition of experimental, communication, cooperation, and evaluation competencies, the development of interest and factual knowledge as well as time management skills. Furthermore, the questionnaire can be used to assess the instructor's guidance (e.g., how good their explanations are and if a good learning atmosphere has been created), the quality of the task instructions, and how well the content from the accompanying lecture is integrated into the course.

Second, instruments have been developed and approaches have been pursued to assess students' acquisition of specific competencies within lab courses to determine the outcome reached. For example, the acquisition of expert-like views on experimental physics in a lab course can be assessed with the E-CLASS instrument (Colorado Learning Attitudes about Science Survey for Experimental Physics; Zwickl et al., 2014), the development of critical thinking skills with the PLIC instrument (Physics Lab Inventory of Critical thinking; Walsh et al., 2019), and the improvement of conceptual understanding with exam tasks (cf. Holmes et al., 2017). Further research on the assessment of experimental skills has for example been conducted by Schreiber (2012) and Bauer (2023).

Both approaches presented before, evaluating the lab course in its entirety and assessing the students' acquisition of competencies, are important and purposeful for evaluating lab concepts in general. However, we argue that they are unsuitable for evaluating individual experimental tasks for physics lab courses e.g., after the design of a new task. The first approach is not expedient because typical lab courses consist of multiple tasks, so the evaluation of a lab course in its entirety only provides limited information for every individual experimental task in this lab course as it rather addresses the overall (average) impression of all experimental tasks in the evaluated lab course. The second approach is inappropriate for evaluating individual experimental tasks because the acquisition of competencies takes place on a larger time scale than the execution of an individual task. So, it can be used to assess the students' development over the course of an entire semester (or study program) but not for evaluating individual experimental tasks. Therefore, a new instrument is needed that focuses on the evaluation of an individual experimental task while considering the two described limitations.

3. Development of the questionnaire

Our goal was the development of a new questionnaire ready to be used for the evaluation of individual experimental tasks for physics lab courses that neither evaluates a lab course in its entirety nor addresses the development of competencies. Instead, it should be a tool to quickly obtain feedback from larger groups of students on a newly developed and firstly implemented experimental task to optimize the experimental task evidence-based for further lab courses.

The primary rationale for the development of such an instrument was the evaluation and improvement of 15 standalone, competence-centered, smartphone-based experimental tasks ready to be used for on-campus and distance learning physics lab courses that were developed in the project "Developing Digital Physics Laboratory Work for Distance Learning" (DigiPhysLab) co-funded by the Erasmus+ program of the European Union (03/21-02/23; cf. Lahme et al., 2022). However, as we realized that a suitable instrument is missing, we hope that this instrument will also serve as a tool for other instructors to evaluate and improve existing or newly developed experimental tasks in their lab courses. We formulated two guiding questions for the development of a new questionnaire:

1. To what extent are the developed experimental tasks from students' point of view suitable for university education?
2. How do students experience working on the experimental task?

Initially, with the focus on evaluating the DigiPhysLab tasks, items to evaluate the experimental tasks were discursively developed and structured within the group of researchers, based on their own experiences, and inspired by already existing instruments and literature about experimental tasks and physics lab courses. The items were initially formulated in English and then translated into German, Finnish, and Croatian. The German, Finnish, and Croatian versions were each communicatively validated with two to three native-speaking physics (teacher training) students (2nd to 5th year of study) to secure that students understand the items as intended. The English version was additionally piloted with a German-speaking physics teacher training student in the 5th year of study and by retranslating the German version back into the English version by an English teacher training student in the 2nd year of study. Feedback from all participating students in all four languages and the results from the re-translation of the German version was used to improve the English version of the questionnaire and its translations consecutively. The final English version can be found in Tab.1 in the appendix, all four versions in a printable, ready-to-be-used mode on our project website (www.jyu.fi/digiphyslab/).

4. Overview of the questionnaire

The questionnaire as displayed in the appendix consists of seven sections with in total of 60 closed items and eight open-text field questions. Section [A] is about the personal information of the participating students to better interpret subsequent responses. Besides demographic data (field and year of study and finished lab courses; items [A1] to [A3]), the general attitude to digital technologies (item [A4]), physics (item [A5]), and lab experiments (item [A6] to [A9]) are addressed. The latter items have been adopted from an addition to the PLIC instrument (Walsh et al., 2019) as done by Pirinen et al. (2023).

Section [B] focuses on the students' perceived learning gains and therefore the efficacy of the experimental task. All items are particularly formulated so that they can be answered for individual experimental tasks e.g., if the concepts in the specific task can be explained to someone else (cf. item [B1]) or if learning goal-related feelings are evoked (e.g., if one feels more confident in conducting experiments, cf. item [B3]). The items in this section address typical learning objectives of physics lab courses (AAPT, 1997; Zwickl et al., 2013; Welzel et al., 1998; Teichmann et al., 2022) like reinforcing concepts, acquiring experimental skills, fostering interest, or acquiring expert-like views on experimental physics without inappropriately assessing the acquisition of competencies in the experimental task. As these items address typical learning objectives, they help to evaluate if the task is suitable for university physics education as intended in the first guiding question of the development of this questionnaire.

Section [C] is about the perceived adequacy of the task referring to the task conception, the designed materials, and the actual implementation (cf. design principles by Lahme et al. (2023b)) and therefore obviously addressing the first guiding question, too. The students are asked to evaluate the quality of the task instructions (items [C1] to [C4], [C9], and [C13]), the relevance and adequacy to their field and level of study (items [C5] and [C8]), the conditions to conduct this task even in distance learning scenarios (items [C6] and [C7], as this was one objective of the DigiPhysLab-project) and the desired circumstances of implementation (items [C10] to [C12]).

Section [D] refers to the students' experience during the task and therefore addresses the second guiding question for the development of this questionnaire. Here, a list of twelve emotions and feelings that one could experience while conducting physics experiments (e.g., frustration, excitement, competency, freedom) is presented to the students and the students can rate to which extent they experienced them. The items [D1] to [D3] are inspired by Schneider et al. (2016) based on the flow theory (Csikszentmihalyi, 2008), the items [D4] to [D12] are adapted from the short version of the *Epistemically-Related Emotion Scales* (Pekrun et al., 2016) and allow insights in the students' perception of the task especially regarding the adequacy of the level of difficulty, the degree of openness, and the achieved engagement during the experiment.

Section [E] focuses on the students' activities rather than the emotions and feelings during the conduction of the experiment and also addresses the second guiding question. 15 statements are provided which contain typical steps of the experimental physics cognition process ranging from the formulation of research questions and hypotheses (items [E1] and [E2]) over the design (items [E3] to [E5]) and data collection (items [E6] to [E8]) to the data analysis (items [E9] to [E13]) and the presentation of findings (items [E14] and [E15]). These items are largely inspired by literature describing typical activities while conducting physics experiments (e.g., Haller, 1999; Millar, 2009; Trinh-Ba, 2016; Tesch & Duit, 2004; Holmes & Lewandowski, 2020). The students rate on the three-point scale "yes – somewhat – no" to which extent they needed to do each activity in the given experimental task. By this, one gains insight into the students' perception of the requirement profile of the task since not every experimental task requires all experimental activities to the same extent. A comparison with the central activities intended by the instructor allows, without an unfeasible measurement of the acquisition of competencies, a reflection of whether the intended experimental skills are fostered.

Section [F] addresses the use of digital technologies in the task and therefore the second guiding question, too. Of course, this section is only applicable if digital technologies are used in the experimental task. However, the use of digital technologies, especially computers for data analysis and visualization, and the use of digital data logging systems (like smartphones, microcontrollers, or computers in remote-controlled labs) is very common in many physics lab courses, so it is reasonable to evaluate if these technologies are from the students' point of view helpful and utilizable for the experimental task. Thus, the students are asked to rate statements about the experienced obstacles while using the technologies (items [F1]), their benefits (items [F2] to [F4] and [F7]), the cost-benefit-ratio of learning how to handle new digital technologies (item [F5]), and the attitude to digital technologies in general (item [F6]). Furthermore, they were asked to describe the impact of digital technologies on the learning process in an open text field (item [F8]).

Section [G] finally is about the students' overall impression of the experimental task and how the experimental task could be further improved. Thus, it addresses both the first and second guiding questions. First (item [G1]), the students are asked to give an overall grade for the task on a scale from 1 (worst) to 10 (best) to get a general impression of the students' perception of the task which allows easy comparisons between different tasks. Consecutively, the students should explain in three open text fields their rating with respect to what they liked, disliked, or would change in the experimental task (items [G2] to [G4]). The responses to these questions are very important to develop further the experimental tasks as they might contain the most explicit feedback.

All in all, the questionnaire collects relevant information on the suitability of the experimental task for physics university education and on the students' experience while working on the task. Through this, students' feedback is gathered to develop further experimental tasks after an (initial) implementation. Depending on the experimental tasks and the instructor's interest, one could add further items to this questionnaire, for example about the students' perception of the factual and disciplinary authenticity of the experimental tasks (cf. Klein, 2016). Based on a sample of $N = 104$ students (three outliers who paused the participation were excluded), fully responding to our questionnaire (and 18 additional closed items on the authenticity of the experimental tasks) takes about $M = 15$ min, $SD = 5$ min. So, using our questionnaire for evaluating experimental tasks is on a time scale comparable to typical course evaluation surveys.

5. Example data and manipulation check

The questionnaire has not been statistically validated yet. Though to showcase as a kind of a manipulation check that the questionnaire reflects differences in the students' perception of different experimental tasks, we present example data of $N = 110$ students for a part of this questionnaire in Fig.1. The questionnaire was used to evaluate undergraduate research projects in a first-year university lecture about mechanics at the University of Göttingen. The students worked in groups of three to five independently over two months on one experimental task that uses smartphones and household items for data collection and computers for data analysis. In total, six different tasks were assigned (each group of students worked on one task) that were originally developed as part of the DigiPhysLab-project but were modified significantly to be suitable for undergraduate research projects. All task instructions followed the same structure and degree of openness but addressed different physics topics. A quick overview of these tasks and the number of students who participated in the questionnaire afterward are displayed in Tab.2 in the appendix (for further information about the tasks see Lahme et al. (2023a) and Lahme et al. (submitted)).

Since the task instructions were designed following the same principles and implemented within the same cohort of students, we would argue that differences in the responses to the questionnaire can mainly be due to differences in the conception of the experimental tasks. In Fig.1, means and standard errors of the students' responses for selected items from section [E]

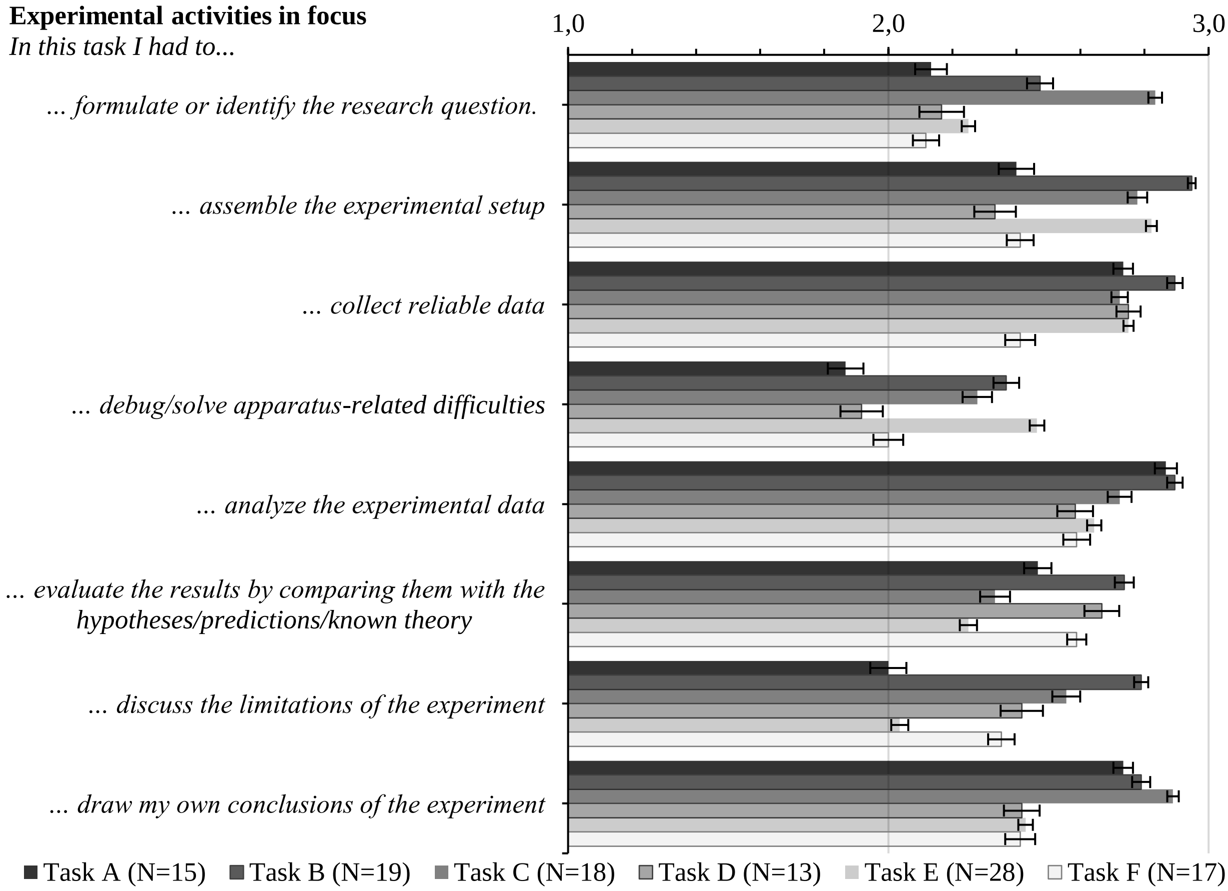

Fig.1: Students' responses (mean and standard error) for selected items of section [E] (experimental activities in focus) for six different experimental tasks characterized in Tab.2. Items were rated on a three-point scale (1 = no, 2 = somewhat, 3 = yes). N indicates the number of responses per task (in total $N = 110$ responses).

(experimental activities in focus) are displayed, broken down by the six tasks. The data demonstrate that the students perceived the relevance of the various experimental activities differently for each task. This is in accordance with the intended task conception.

For example, for all but tasks B and C the research question was given in the task instructions while in task B just the main research question was given, and sub-questions needed to be formulated and in task C students needed to come up with their own research question according to a given overall goal. These differences are reflected in the questionnaire responses as the need for identifying/formulating their own research question was rated significantly higher for task C than tasks A, D, E, and F, with task B in between.

Similarly, students' perception is largely in line with the intended task conception for assembling the experimental setup that is demanded rather in tasks C and E than in A and D. Just the high rating for task B is not as expected but can be explained with reported difficulties to procure suitable household items for the experiment (e.g., falling bodies and scales) and not a higher demand in setting up the experiment.

The students perceived (in accordance with what they were asked in the task instructions) the need for collecting reliable data as similarly high for all tasks except task F. However, this task is the only one in which data should primarily be analyzed qualitatively, so that is exactly as intended.

The responses to the other items can be similarly discussed, so the data provide anecdotal evidence that the questionnaire enables students to provide differentiated feedback on experimental tasks. This feedback can serve instructors as a basis for reflecting on their implemented experimental tasks and developing them further based on the students' feedback, especially in the case that the students' responses do not go along with the instructor's intentions.

6. Summary and outlook

Students' feedback on experimental tasks in physics lab courses can help to reflect their quality and to develop them further. As already existing instruments and approaches focus either on the evaluation of lab courses in their entirety or on the students' acquisition of specific competencies, we developed a new questionnaire that provides information about the students' perception of the task suitability for university education and about the students' experience while working on the experimental task. The responses on the efficacy/perceived learning gains, the adequacy of the task, the students' experience during the task, the experimental activities in focus, the use of digital technologies in the task, and the overall impression of the task help instructors to reflect and re-design their experimental tasks. Example data show that students indeed respond differently to the questionnaire for different experimental tasks and that their responses are in accordance with the intended task conception.

In the future, the questionnaire needs to undergo a more advanced statistical analysis, for example, regarding scale reliability. For this, the instrument should be largely used in different lab courses and for different tasks to assemble variable data. Moreover, the validity should not only be checked with students as done in this work but also with experts.

7. References

American Association of Physics Teachers (1997). Goals of the Introductory Physics laboratory. *The Physics Teacher*, *35*(9), 546–548. https://aapt.scitation.org/doi/pdf/10.1119/1.2344803

Bauer, A. B. (2023). *Experimentelle Kompetenz Physikstudierender: Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden* [Thesis]. Universität Paderborn, Paderborn. https://doi.org/10.17619/UNIPB/1-1652

Costello, P. J. M. (2003). *Action research*. *Continuum Research Methods*. Continuum.

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience*. Harper Collins.

Haller, K. (1999). *Über den Zusammenhang von Handlungen und Zielen: Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*. Logos.

Helmke, A. (2002). Kommentar: Unterrichtsqualität und Unterrichtsklima: Perspektiven und Sackgassen. *Unterrichtswissenschaft*, *30*(3), 261-277. https://www.pedocs.de/volltexte/2013/7689/pdf/UnterWiss_2002_3_Helmke_Kommentar.pdf

Holmes, N. G., & Lewandowski, H. J. (2020). Investigating the landscape of physics laboratory instruction across North America. *Physical Review Physics Education Research*, *16*(2), 020162. https://doi.org/10.1103/PhysRevPhysEduc Res.16.020162

Holmes, N. G., Olsen, J., Thomas, J. L., & Wieman, C. E. (2017). Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content. *Physical Review Physics Education Research*, *13*(1), 010129. https://journals.aps.org/prper/abstract/10.1103/PhysRevPhysEducRes.13.010129

Klein, P. (2016). *Konzeption und Untersuchung videobasierter Aufgaben im Rahmen vorlesungsbegleitender Übungen zur Experimentalphysik (Mechanik)* [Thesis]. Technische Universität Kaiserslautern, Kaiserslautern. https://www.uni-goettingen.de/de/document/download/8f76f0f86ffc48a34292baaaddbaead8.pdf/Dissertation%20KLEIN%20Pascal.pdf

Lahme, S. Z., Klein, P., Lehtinen, A., Müller, A., Pirinen, P., Sušac, A., & Tomrlin, B. (2022). DigiPhysLab: Digital Physics Laboratory Work for Distance Learning. *PhyDid B – Didaktik der Physik – Beiträge zur DPG-Frühjahrstagung virtuell 2022*, 383-390. https://ojs.dpg-physik.de/index.php/phydid-b/article/view/1250/1503

Lahme, S. Z., Fipp, M., Müller, A. & Klein, P. (submitted). Offene Projektaufgaben mit Smartphone-Experimenten für die Studieneingangsphase Physik. *PhyDid B – Didaktik der Physik – Beiträge zur DPG-Frühjahrstagung Hannover 2023* [in this proceedings].

Lahme, S. Z., Müller, A., & Klein, P. (2023a). Lehrveranstaltungsverbindende Experimentieraufgaben im Physikstudium. *In v. Vorst, H. (Hrsg.). Lernen, lehren und forschen in einer digital geprägten Welt, Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Aachen 2022, 43*, 663-666. https://gdcp-ev.de/wp-content/uploads/securepdfs/2023/05/P020_Lahme.pdf

Lahme, S. Z., Pirinen, P., Rončević, L., Lehtinen, A., Sušac, A., Müller, A., Klein, P. (2023b). *A framework for designing experimental tasks in contemporary physics lab courses*. Preprint on arXiv, submitted to the proceedings of GIREP conference Ljubljana 2022. http://dx.doi.org/10.48550/arXiv.2302.14464

Millar, R. (2009). *Analysing practical activities to assess and improve effectiveness: The Practical Activity Analysis Inventory (PAAI)*. York. Centre for Innovation and Research in Science Education, University of York. https://www.rsc.org/cpd/teachers/content/filerepository/frg/pdf/Researchby-Millar.pdf

Müller, A., & Brown, A. (2022). An evidence-based approach to tasks in science education: Meta analytical and other quantitative results. *Progress in Science Education*, *5*(1), 6-32. https://doi.org/10.25321/prise.2022.1275

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268-1276. https://doi.org/10.1080/02699931.2016.1204989

Pirinen, P., Lehtinen, A., & Holmes, N. G. (2023). Impact of traditional physics lab instruction on students' critical thinking skills in a Finnish context. *European Journal of Physics*, 44, 035702. https://doi.org/10.1088/1361-6404/acc143

Rehfeldt, D. (2017). *Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika*. Logos.

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., Bruner, J., Moeller, J., Linnansaari, J., Juuti, K., & Viljaranta, J. (2016). Investigating optimal learning moments in U.S. and Finnish science classes. *Journal of Research in Science Teaching*, *53*(3), 400–421. https://doi.org/10.1002/tea.21306

Schreiber, N. (2012). *Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*. Logos.

Teichmann, E., Lewandowski, H. J., & Alemani, M. (2022). Investigating students' views of experimental physics in German laboratory classes. *Physical Review Physics Education Research*, *18*(1), 010135. https://doi.org/10.1103/PhysRevPhysEduc-Res.18.010135

Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht: Ergebnisse einer Videostudie. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *10*, 51–69.

Trình-Bá, T. (2016). *Development of a course on integrating ICT into inquiry-based science education* [Thesis]. Vrije Universiteit Amsterdam, Amsterdam. https://research.vu.nl/ws/portalfiles/portal/42159760/complete+dissertation.pdf

Walsh, C., Quinn, K. N., Wieman, C., & Holmes, N. G. (2019). Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking. *Physical Review Physics Education Research*, *15*(1), 010135. https://doi.org/10.1103/PhysRevPhysEduc-Res.15.010135

Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, A., Robinault, K., & Aufschnaiter, S. von (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden: Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, *4*(1), 29–44.

Zwickl, B. M., Finkelstein, N., & Lewandowski, H. J. (2013). The process of transforming an advanced lab course: Goals, curriculum, and assessments. *American Journal of Physics*, *81*(1), 63–70. https://doi.org/10.1119/1.4768890

Zwickl, B. M., Hirokawa, T., Finkelstein, N., & Lewandowski, H. J. (2014). Epistemology and expectations survey about experimental physics: Development and initial results. *Physical Review Special Topics Physics Education Research*, *10*(1), 010120. https://doi.org/10.1103/PhysRevSTPER.10.010120

Author contributions*

Pascal Klein: Conceptualization (supporting), Funding Acquisition (supporting), Supervision (equal), Writing – review & editing (supporting). Simon Z. Lahme: Conceptualization (lead), Visualization, Writing – original draft, Writing – review & editing (lead). Antti Lehtinen: Conceptualization (supporting), Funding acquisition (lead), Project administration, Supervision (equal), Writing – review & editing (supporting). Andreas Müller: Supervision (equal), Writing – review & editing (supporting). Pekka Pirinen: Conceptualization (lead), Writing – review & editing (supporting). Lucija Rončević: Conceptualization (lead), Writing – review & editing (supporting). Ana Sušac: Conceptualization (supporting), Funding Acquisition (supporting), Supervision (equal), Writing – review & editing (supporting).

* According to CREDIT (CRediT Contributor Roles Taxonomy), https://credit.niso.org

Funding

We are grateful for the financial support by the Erasmus+ program of the European Union (G.A.-No.: 2020-1-FI01-KA226-HE-092531).

Appendix

Tab.1: The full (English) questionnaire for evaluating experimental tasks in physics laboratory courses. The German, Finnish, and Croatian versions of this questionnaire (as well as the English one) can be found in a printable, ready-to-be-used version at https://www.jyu.fi/digiphyslab/.

Item(s)	Scale
[A] Personal information	
[A1] What is your major field of study?	open text field
[A2] What is your year of study?	
[A3] Please list all lab courses you have finished during your studies.	
[A4] In general, I am interested in digital technologies.	strongly disagree – disagree – I do not disagree or agree – agree – strongly agree
[A5] In general, I am interested in physics.	
In general, where would you put doing lab experiments on the following scales of opposites?	
[A6] Boring - Interesting	1 (left adjective) – 2 – 3 – 4 – 5 (right adjective)
[A7] Useless – Useful	
[A8] Hard – Easy	
[A9] Stressful – Fun	
[B] Efficacy/perceived learning gains	
After completing the task…	strongly disagree – disagree – I do not disagree or agree – agree – strongly agree
[B1] … I could explain the physical concepts in this task to someone else.	
[B2] … I could explain what I have done in this task to someone else.	
[B3] … I feel more confident in conducting physics experiments.	
[B4] … I am more interested in conducting physics experiments.	
[B5] … I have a better insight into what research in experimental physics looks like.	
[B6] … I feel like I learned something new.	
[B7] … I feel more confident in using digital technologies in the lab.	
[C] Adequacy of the task	
[C1] The learning objectives of the task were clear to me.	strongly disagree – disagree – I do not disagree or agree – agree – strongly agree
[C2] The task instructions were easy to understand.	
[C3] The task instructions had a clear layout.	
[C4] Instructions on how to use the digital technologies in this task were sufficient for me.	
[C5] It is clear to me how this task is related to my field of study.	
[C6] I have the conditions (e.g., necessary equipment) to conduct this experimental task at home.	
[C7] I feel confident that I could do the experiment on my own at home.	
[C8] This experimental task was *too easy/adequately challenging/too difficult* for my level of study.	The italic answers are selectable.
[C9] Task instructions and supportive materials were *too detailed/sufficient/ insufficient*.	
[C10] How much time would you need to complete this experimental task without any pressure?	open text field
[C11] I would prefer to do this task *on campus/synchronously (e.g., during a web conference) at home/asynchronously (e.g., support only via e-mail) at home/no preference*.	The italic answers are selectable.
[C12] For this task, I would prefer to *work alone/in pairs/in small groups*.	
[C13] Which task instructions were confusing?	open text field
[D] Students' experience during the task	
[D1] During this task I felt skilled at what I was doing.	strongly disagree – disagree – I do not disagree or agree – agree – strongly agree
[D2] During this task, I was interested in what I was doing.	
[D3] I understood this task as a challenge.	
[D4] During this task, I felt surprised.	
[D5] During this task, I felt curious.	
[D6] During this task, I felt excited.	
[D7] During this task, I felt confused.	
[D8] During this task, I felt anxious.	
[D9] During this task, I felt frustrated.	

[D10] During this task, I felt bored.	
[D11] I had opportunities to use my creativity in designing and conducting experiments.	
[D12] I had opportunities to make my own decisions about the experiment.	
[E] Experimental activities in focus	
In this task, I had to…	
[E1] … formulate or identify the research question.	no – somewhat – yes
[E2] … formulate my own hypothesis.	
[E3] … assemble the experimental setup.	
[E4] … decide what physical quantities I need to measure in the experiment.	
[E5] … decide how to measure physical quantities in the experiment.	
[E6] … collect reliable data.	
[E7] … debug/solve apparatus-related difficulties.	
[E8] … document the experimental process.	
[E9] … analyze the experimental data.	
[E10] … determine the uncertainty of the experimental data.	
[E11] … evaluate the results by comparing them with the hypotheses/predictions/known theory.	
[E12] … discuss the limitations of the experiment.	
[E13] … draw my own conclusions of the experiment.	
[E14] … use different representations for data visualization (graphs, tables, …).	
[E15] … present and discuss the results of the experiment using scientific terminology (e.g., in a lab report or an oral presentation).	
[F] Use of digital technology in the task	
[F1] The use of digital technologies made this task difficult.	strongly disagree – disagree – I do not disagree or agree – agree – strongly agree
[F2] Digital technologies made this task interesting.	
[F3] The digital technologies helped me to develop further my experimental skills.	
[F4] The digital technologies helped me get a better understanding of the physical concepts.	
[F5] The effort to learn how to use digital technologies in this task was worthwhile.	
[F6] I prefer to use standard lab equipment instead of digital technologies like smartphones/simulations.	
[F7] Digital technologies made performing the task easier.	
[F8] In what way did the use of digital technologies impact your learning process?	open text field
[G] Final open questions	
[G1] Rate this task based on your overall impression (regardless of your own performance) on a scale of 1 (worst) to 10 (best)	1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10
[G2] What did you like about the task? And why?	open text field
[G3] What did you dislike about the task? And why?	
[G4] What would you change in this task? And why?	

Tab.2: Topic/goal and characterization of the six experimental tasks for which example data are displayed. The task documents are available at https://doi.org/10.57961/49zr-w490.

Task	Topic/goal	Characteristics
A Slamming door	Modeling frictional effects of a slamming door	Focus on data analysis (fitting data & testing known models)
B Paper parachute	Video analysis of the velocity dependency of air friction in free fall	Own sub-questions needed to be formulated, and video analysis for answering the research questions
C Sensor analysis	Comparing the precision of acceleration sensors of different smartphones	An own experiment (research question, setup, etc.) needed to be developed to compare the sensors.
D Elevator oscillations	Analysis of the relationship between the period duration and rope length of an oscillating elevator cabin	No setup is needed, focus more on preparing and analyzing data
E Rolling smartphone	Analysis of the relationship between the angle of a declined plane and the parameters of a rolling tin on it	Own setup needed to be assembled, focus on data analysis
F Rotating smartphone	Analysis of parameters of a free rotating smartphone over time	More qualitative analysis & interpretation of graphs investigating which rotations are (un-)stable