European Journal of Physics

Impact of traditional physics lab instruction on students' critical thinking skills in a Finnish context

Impact of traditional physics lab instruction on students' critical thinking skills in a Finnish context

P. Pirinen[1], **A. Lehtinen**[1,2], **N. G. Holmes**[3]

[1]Department of Physics, P.O. Box 35, 40014 University of Jyväskylä, Finland
[2]Department of Teacher Education, P.O. Box 35, 40014 University of Jyväskylä, Finland
[3]Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

E-mail: pekka.a.pirinen@jyu.fi

Abstract. Recent studies have given incentive to physics departments around the world to revise the learning goals of their lab courses to emphasize experimentation skills over reinforcing lecture content. Evaluation instruments have been developed to measure the achievement of learning goals, and one such instrument is the PLIC (Physics Lab Inventory of Critical thinking). The PLIC measures respondents' ability to evaluate models, evaluate methods, and to suggest next steps for an investigation. In the present work we give consideration to the validity of our Finnish translation of the PLIC and we show results from our baseline study of first-year introductory lab courses with labs mixing content- and skills-related learning goals. We observed no statistically significant change in students' critical thinking skills over the period of our study.

1. Introduction

Courses in experimental physics, often referred to as lab courses, are an important part of a university physics curriculum. A multitude of learning goals have been traditionally assigned to lab courses, such as reinforcing lecture course content, linking theory and practice, developing experimental skills, learning about the nature of science and practicing communication and collaboration skills [1–3]. In recent years there has been increasing concern that some of these learning goals of laboratory courses are very often not fulfilled. In particular, there has been difficulty in finding any impact of laboratory instruction to learning and understanding physics concepts or reinforcing theory content [4–10]. It has therefore been suggested to instead focus explicitly on teaching experimentation skills, for example setting a research question, experimental design, collecting and analyzing data, evaluating data and methods, and communication skills. This approach has measurable benefits in students' critical thinking skills [9, 10] as well as developing more expertlike attitudes and beliefs about experimental physics

[9–13], while maintaining a similar level of content-related learning outcomes as students' participating in lab courses aiming specifically to reinforce lecture content [5,6,9].

Even though the need for authentic inquiry and actual scientific discovery in physics education has been recognized essentially from the get-go [14,15], it is fairly common that contemporary experimental tasks follow a cookbook-style instruction, where students have very limited decision-making opportunities [14]. It is suggested that this approach would require the perspective of a seasoned expert to recognize the choices made behind the complexity of the experimental setup, and students lacking this perspective will resort to just following instructions, without the satisfaction of discovery [16].

In combination with highly detailed step-by-step instruction, laboratory tasks typically involve replicating some known result or confirmation of a well-established theory. Such confirmational activities have been shown to in some cases lead to questionable research practices such as manipulating data and other attempts at moving the goalposts so that the desired results could be apparently concluded from the performed experiment [17]. It is therefore highly important to revise also the structure of experimental tasks included in physics lab courses.

The Deparment of Physics at the University of Jyväskylä is preparing for a full reform of the laboratory studies in the bachelor level. The lab courses studied in the present baseline work contained a mix of learning goals related to reinforcing and demonstrating lecture content, experimentation skills, and communication skills. Lab assignments within the courses were mostly traditional highly structured tasks related directly to lecture content on simultaneously running theory courses. Guided by the evidence presented above, a need for changes is recognized, and new courses in experimental physics are currently being planned starting from a set of skills-based learning outcomes to be developed during the full studies. The new introductory lab courses will be piloted gradually during the next two years, and the pilots will be researched and courses developed further. The new curriculum will be implemented in full in fall 2024.

Our aim is to study the transformation of a bachelor-level curriculum with mostly structured traditional labs mixing concepts-based and skills-based learning goals to a new set of courses in experimental physics which explicitly aim to teach experimentation skills while providing possibilities and support for student agency. In this article we report results of our baseline study over one academic year before any of the planned changes take place. To quantify the effect of our lab courses, we use the evaluation instrument Physics Lab Inventory of Critical thinking (PLIC) [18]. Our research questions for this study are

RQ1: Is the Finnish translation of the PLIC instrument valid for use in the Finnish context?

RQ2: How do students' critical thinking abilities develop within their first year of studying physics in a lab curriculum with both skills-based and concepts-based learning goals, and how do students' attitudes toward experimental physics evolve in the same period?

In Section 2 we describe the methods used in our study. In Section 3 we discuss our findings. In Section 4 we make concluding remarks of the baseline study for the upcoming curriculum change, and consider limitations in the present analysis.

2. Methods

2.1. The evaluation instrument PLIC

To measure the effect of the lab courses in student's critical thinking skills, we used the instrument Physics Lab Inventory of Critical thinking (PLIC) [18] developed by researchers of the Cornell Physics Education Research Lab. In the PLIC the respondents are given two scenarios of physicists conducting an experiment to test a mathematical model for the period of a spring-and-mass system. The two groups have designed different experiments to test the model, and the respondents answer questions which test their critical thinking skills related to evaluating models, evaluating methods, and suggesting next steps for the investigation [18].

In the current version of the PLIC there are 8 multiple-response items which are scored between 0 and 1 points via weights based on answers obtained from expert physicists. The original scoring scheme is described in Ref. [18]. The scoring scheme of the PLIC was updated during our data collection, and one of the survey items (Q4B) was changed between our first presurvey and postsurvey. We decided to leave that question out of our total-score analyses, so that we have 7 scored questions in the survey, and the maximum total score obtainable was therefore 7.

There are three distinct types of scored questions in the PLIC. The first type is an "Evaluate data" question (Q1B, Q2B, Q3B), which asks the respondent to provide support for their reasoning for whether presented data agree with the given model. "Next steps" questions (Q1E, Q2E, Q3E) ask the respondents what they think the physicists should do next based on their data. The"Evaluate methods" question (Q3D) is about comparing different analysis methods: why group 2 should use one fit to their data over another.

Via additional questions incorporated in the PLIC survey, we collected data on what students think about experimental physics. This involved rating experimental work between opposite adjectives (e.g. fun–scary, interesting–boring), and a Likert-scale evaluation of statements about experimental work (e.g. I am confident in my ability to analyze data).

2.2. Finnish translation of the PLIC and validation

The PLIC was translated into Finnish by the first two authors. The first version was translated by P. Pirinen, subsequently revised by A. Lehtinen, and then discussed until we reached consensus on each survey item. To mitigate the effects of the translation we performed a further communicative validation with three people (students pursuing a master's degree or a PhD). The interviewees went through each item in the translated survey and described how they understood it. Any possible double meanings and discrepancies from the original version were recorded, and in unclear cases a direct comparison with the original English version of the PLIC was made with the interviewees. Our translated Finnish version of the PLIC is available online [19].

2.3. Data collection

We collected data from students enrolled in the first introductory lab course at the University of Jyväskylä. We did a PLIC presurvey in August 2021 at the start of the academic year. We got responses from 58 students. The PLIC postsurvey was conducted in April 2022, with 26 respondents. Of these, 24 students responded to both the presurvey and the postsurvey. Unless explicitly specified, any following pre-post-survey comparisons will be based on this group that responded to both, with $N^{\mathrm{JYU}}_{\mathrm{both}} = 24$. In the beginning of the fall semester in August 2022, before any lab instruction, we again collected pre-survey data from the first introductory lab course, yielding 69 responses boosting the number of total pre-survey responses to $N^{\mathrm{JYU}}_{\mathrm{pre}} = 127$. This number can be reflected against the total number of roughly 230–260 students yearly (in 2018–2022) starting in an undergraduate degree program in physics in Finland [20].

During the period between the pre-survey and post-survey students participated in at least one lab course. Most lab activities within the courses were highly structured so that students were given direct instructions on what to measure, how to measure it, and how to present and analyze their data. The labs were related to simultaneously running theory courses, and the tasks demonstrate a phenomenon related to lecture content, such as projectile motion or the effect of humidity on air density. Several lab assignments also have confirmatory goals (e.g., determine the charge-to-mass ratio of the electron), either explicitly or implicitly, so that it is expected that students observe the phenomenon as predicted by the presented theory and models.

The lab work on the course(s) was not bound to a specific schedule, and students were able to reserve time slots for each lab task freely. The courses thus do not have a strict end date, and even though it was recommended to complete all lab assignments by the end of the spring semester, this did not happen in most cases. Moreover, while every participant in our survey was enrolled in the first introductory lab course, not everyone enrolled in the other available courses later. Therefore, there is significant variance in the number of lab activities completed by students during the two semesters. A lab activity was typically a 4-hour experimentation in the student lab, with either a shorter informal reporting included or a full written report done at home afterwards. Out of the 17 lab activities available to the students, the minimum number of activities completed by respondents was 4, the maximum was 16, and the median was 9. However, no correlation was observed between the number of lab activities done and the difference in pre- and post-survey scores (Pearson's $r = -0.059$). In addition to the lab activities, students participated in lectures and did exercises. Thus we consider that all students in the matched pre-post dataset have received enough instruction in experimental physics to be included in the study, despite the variance in the number of completed lab activities.

3. Results and discussion

3.1. Validation of the Finnish translation of the PLIC

We performed a communicative validation of the Finnish translation of the PLIC as described in Section 2. We received a few possible discrepancies between the Finnish translation and the original English version of the PLIC from each of the three interviews. We considered changes to our translation based on the following two criteria: 1) an item pointed out by a single interviewee was clearly different between the Finnish and English versions of the PLIC or 2) the same potential discrepancy was pointed out by at least 2 of the 3 interviewees. As a result of the communicative validation, no clear discrepancies between the Finnish and English versions were found, and only the translation of an answer option "The data neither support nor refute the model", was made clearer for each question it appears in.

To further examine the validity of the Finnish translation of the PLIC, we compare our pre-survey data collected over two years, referred to later as JYU data, to data collected from a similar demographic of introductory lab course participants in North America, referred to as NA data. The NA data consists of $N_{\mathrm{pre}}^{\mathrm{NA}} = 445$ pre-survey responses. The average total score (not including question Q4B as discussed earlier) in the JYU data is 2.63 ± 0.08 and in the NA data it is 2.62 ± 0.05. On the overall level the survey scores seem consistent for students at the beginning of their university physics studies whether they study in Finland or in North America. This gives us confidence that the translated version of the PLIC can be used to measure students' critical thinking abilities in a similar way as the original does.

In Figure 1 we show the average points per evaluated question in the PLIC in the pre-survey JYU and NA datasets. The questionwise scores for the two datasets seem reasonably similar for students at a similar academic level in that the difference in any single question does not appear striking. We show the average score per question within the distinct question types of "Evaluate data" questions (Q1B, Q2B, Q3B), "Next steps" questions (Q1E, Q2E, Q3E), and "Evaluate methods" question (Q3D) between the JYU and NA datasets in Figure 2. There is a statistically significant difference in the average scores of the JYU data and NA data for each question type (independent samples t test, Holm-Bonferroni corrected, $\alpha < 0.05$) with small effect size (Cohen's $d = 0.40, 0.35, 0.22$ for the B, D, and E type questions, respectively). On average, students in the JYU scored slightly higher in the "Evaluate data (B)" type questions than students in NA, while in the "Evaluate methods (D)" and "Next steps (E)" questions the NA students scored higher. The scores can be different due to a multitude of reasons which cannot be pinned down from this data alone. The differences observed here are small, but whether the critical thinking skills measured by each type of question in the PLIC are universal is an interesting question which could be explored from a wider international dataset. See more discussion about this in Section 4

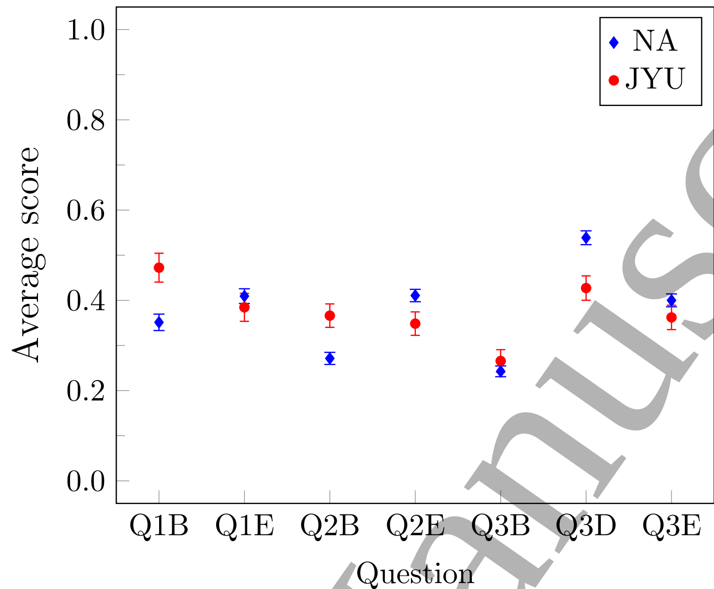

Figure 1: Mean scores and standard errors for each evaluated question in the PLIC for the University of Jyväskylä (JYU) pre-test data (red), and the reference pre-test dataset from North America (blue).

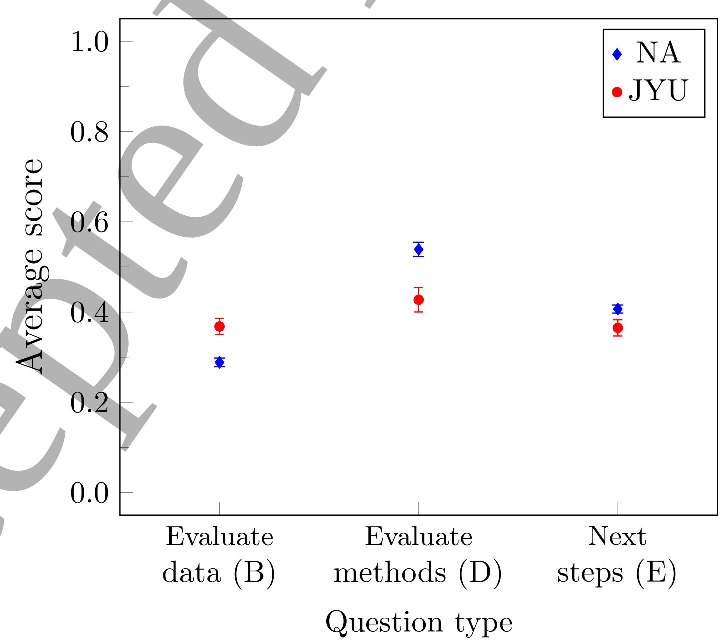

Figure 2: Mean scores and standard errors for questions within each question type of the PLIC for the University of Jyväskylä (JYU) pre-test data (red), and the reference pre-test dataset from North America (blue).

3.2. Development of students' critical thinking skills

In our matched PLIC dataset ($N = 24$), the average total scores (out of maximum of 7) were 2.44 ± 0.17 in the pre-test and 2.68 ± 0.16 in the post-test. The difference is not statistically significant (paired t-test, $p = 0.279$). The average pre- and post-test scores per question are shown in Figure 3. Comparing means for each question, we found no statistically significant pre-post-test differences in any of the 7 questions (Wilcoxon signed-rank test, Holm-Bonferroni corrected, $\alpha < 0.05$).

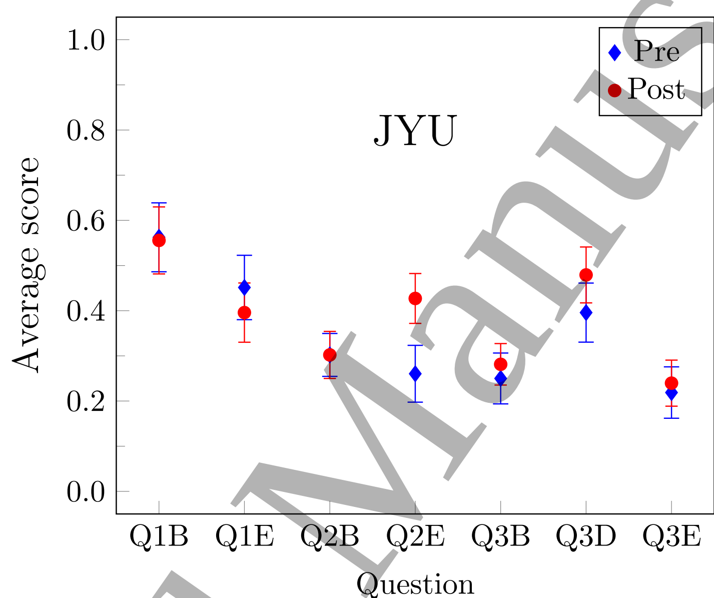

Figure 3: Mean scores and standard errors for each evaluated question in the PLIC for the University of Jyväskylä pre-test in August 2021 (blue), and post-test in April 2022 (red). The number of students participating in both surveys was $N = 24$, and means are shown only for this group.

Based on the current data, we are unable to claim that introductory physics lab courses at the University of Jyväskylä at the time of this study lead to any improvements in students' critical thinking skills as measured by the PLIC. This is consistent with results obtained in Ref. [10] for courses that focus on reinforcing physics concepts rather than teaching experimental skills.

In table 1 we show our pre- and post-test results for questions concerning attitudes and beliefs about experimental physics. For these questions outside the main PLIC survey the number of respondents was $N = 23$. There is a statistically significant difference (paired t-test, Holm-Bonferroni corrected, $\alpha < 0.05$) between the pre-survey and post-survey in only one item, in which students rated doing lab experiments between opposite adjectives Useful and Useless, with medium effect size (Cohen's $d = 0.66$). Students felt like doing lab experiments is not as useful after two semesters containing lab instruction as in the beginning of the first semester. It has been established before that lab courses which focus on reinforcing physics concepts and do not offer enough decision-making opportunities cause students' attitudes and beliefs about experimental physics to shift to a less expertlike direction [9–11, 13]. The decrease in the perceived usefulness of doing lab experiments observed in the present work could be a manifestation of the same phenomenon.

Table 1: Pre- and postsurvey results ($N = 23$) for questions regarding students' thoughts about experimental physics. In columns 2 and 3 we list the mean and standard error of the students' responses to each item of the pre- and postsurvey, respectively.

Where would you put doing lab experiments on the following scales between two opposite adjectives? (5 point scale)		
Item	Avg (pre)	Avg (post)
Interesting 1 – 5 Boring	1.83 ± 0.13	2.26 ± 0.22
Useful 1 – 5 Useless	1.35 ± 0.11	2.00 ± 0.19
Easy 1 – 5 Hard	3.48 ± 0.19	3.65 ± 0.14
Fun 1 – 5 Scary	2.70 ± 0.16	2.48 ± 0.17
Please indicate how well you agree with the following statements: (1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree)		
Item	Avg (pre)	Avg (post)
I feel confident analyzing data	3.30 ± 0.19	3.30 ± 0.23
I feel confident doing experiments in lab	3.00 ± 0.22	3.04 ± 0.27
Conducting experiments is important to the understanding of science	4.43 ± 0.20	4.43 ± 0.20
The main purpose of experiment is to verify theory	3.91 ± 0.23	3.87 ± 0.20

4. Limitations and Conclusions

In this work we have examined the validity of our Finnish translation of the PLIC survey, and presented findings of our baseline study of first-year introductory lab courses at the University of Jyväskylä (JYU). Based on a communicative validation and comparisons of pre-survey data collected at JYU to data collected from North America, it appears that our translation of the PLIC can be used to measure critical thinking skills in a similar way as the original English version. From data collected from introductory lab course participants at JYU between August 2021 and April 2022, we found no statistically significant changes in the students' critical thinking skills as measured by the PLIC. As a secondary conclusion, we found a small statistically significant reduction in the perceived usefulness of doing lab experiments. Both of these findings are typical to courses that focus explicitly in reinforcing lecture content and offer limited decision-making opportunities to students [9, 10].

The baseline data presented here will provide a benchmark to which we can compare the effects of renewed lab instruction at JYU in subsequent years. Further research in the Finnish context is needed to uncover the possible development of students' critical thinking skills while participating in an explicitly skills-based lab course.

Our sample size of $N = 24$ for students responding to both the pre-survey and post-survey is quite small, which limits the conclusions of the present work. This is true especially for the questions regarding attitudes and thoughts about experimental physics ($N = 23$), where fluctuation in a single student's answers is expected not only due to a long-term shift in attitudes due to lab instruction, but also due to short-term effects like current mood when answering the survey.

As an additional note, the PLIC has been developed using responses from students of North American institutions and the scoring scheme is based on views of experts working in North American institutions. There exists a possibility that the survey is not directly and exactly transferrable to a different part of the world while measuring critical thinking abilities in the exact same way. However, we expect the overall effect of this to be small for the purposes of the present study. Still, a question could be raised of how universal the critical thinking skills as measured by the PLIC are. Further international comparison of data from introductory lab courses in several countries from different parts of the world is needed to find out whether there are universal trends in PLIC responses or if there exist some regional differences.

Acknowledgments

This work was supported by the Erasmus+ 2020-1-FI01-KA226-HE-092531 project Developing Digital Physics Laboratory Work for Distance Learning – DigiPhysLab.

Ethical statement

Per the research ethics requirements of the University of Jyväskylä the research did not require approval from the ethics committee.

The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. All participants gave written informed consent to participate in the study.

References

[1] Trumper R 2003 *Science and Education* **12** 645–670 ISSN 09267220 URL `http://link.springer.com/10.1023/A:1025692409001`

[2] Zwickl B M, Finkelstein N and Lewandowski H J 2013 *American Journal of Physics* **81** 63–70 (*Preprint* https://doi.org/10.1119/1.4768890) URL https://doi.org/10.1119/1.4768890

[3] MacIsaac D 2015 *The Physics Teacher* **53** 253–253 (*Preprint* https://doi.org/10.1119/1.4914580) URL https://doi.org/10.1119/1.4914580

[4] Séré M G 2002 *Science Education* **86** 624–644 (*Preprint* https://onlinelibrary.wiley.com/doi/pdf/10.1002/sce.10040) URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.10040

[5] Etkina E, Van Heuvelen A, Karelina A, Ruibal-Villasenor M and Rosengrant D 2007 *AIP Conference Proceedings* **951** 88–91 (*Preprint* https://aip.scitation.org/doi/pdf/10.1063/1.2820955) URL https://aip.scitation.org/doi/abs/10.1063/1.2820955

[6] Etkina E, Karelina A, Ruibal-Villasenor M, Rosengrant D, Jordan R and Hmelo-Silver C E 2010 *Journal of the Learning Sciences* **19** 54–98 (*Preprint* https://doi.org/10.1080/10508400903452876) URL https://doi.org/10.1080/10508400903452876

[7] Wieman C and Holmes N G 2015 *American Journal of Physics* **83** 972–978 ISSN 0002-9505, 1943-2909 URL http://aapt.scitation.org/doi/10.1119/1.4931717

[8] Holmes N, Olsen J, Thomas J L and Wieman C E 2017 *Physical Review Physics Education Research* **13** 010129 ISSN 2469-9896 URL http://link.aps.org/doi/10.1103/PhysRevPhysEducRes.13.010129

[9] Smith E M, Stein M M, Walsh C and Holmes N 2020 *Physical Review X* **10** 011029 ISSN 2160-3308 URL https://link.aps.org/doi/10.1103/PhysRevX.10.011029

[10] Walsh C, Lewandowski H and Holmes N 2022 *Physical Review Physics Education Research* **18** 010128 ISSN 2469-9896 URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.18.010128

[11] Wilcox B R and Lewandowski H J 2017 *Phys. Rev. Phys. Educ. Res.* **13**(1) 010108 URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.13.010108

[12] Kontro I, Heino O, Hendolin I and Galambosi S 2018 *European Journal of Physics* **39** 025702 URL https://doi.org/10.1088/1361-6404/aa9364

[13] Henderson R, Funkhouser K and Caballero M D 2020 A longitudinal exploration of students' beliefs about experimental physics *2019 Physics Education Research Conference Proceedings* (Provo, UT: American Association of Physics Teachers) URL https://www.compadre.org/per/items/detail.cfm?ID=15278

[14] Hofstein A and Lunetta V N 2004 *Science Education* **88** 28–54 (*Preprint* https://onlinelibrary.wiley.com/doi/pdf/10.1002/sce.10106) URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.10106

[15] Otero V K and Meltzer D E 2016 *The Physics Teacher* **54** 523–527 (*Preprint* https://doi.org/10.1119/1.4967888) URL https://doi.org/10.1119/1.4967888

[16] Wieman C 2015 *The Physics Teacher* **53** 349–351 (*Preprint* https://doi.org/10.1119/1.4928349) URL https://doi.org/10.1119/1.4928349

[17] Smith E M, Stein M M and Holmes N G 2020 *Phys. Rev. Phys. Educ. Res.* **16**(1) 010113 URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.16.010113

[18] Walsh C, Quinn K N, Wieman C and Holmes N 2019 *Physical Review Physics Education Research* **15** 010135 ISSN 2469-9896 URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.15.010135

[19] The PLIC and the Finnish translation of the PLIC are available on PhysPort https://www.physport.org/assessments/assessment.cfm?A=PLIC accessed: 06.02.2023

[20] Vipunen – Education Statistics Finland https://vipunen.fi/en-gb/ accessed: 03.02.2023