Review Article

Assessment designs of instructional labs: A literature review and a design model

Laura Ketonen,[1] Antti Lehtinen,[2] and Pekka Koskinen[3]

[1]*Department of Teacher Education/Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, FI-40014, Finland*

[2]*Department of Physics/Department of Teacher Education, University of Jyväskylä, Jyväskylä, FI-40014, Finland*

[3]*Department of Physics, University of Jyväskylä, Jyväskylä, FI-40014, Finland*

(Received 12 January 2023; accepted 6 June 2023; published 17 July 2023)

[This paper is part of the Focused Collection on Instructional labs: Improving traditions and new directions.] In recent years, physics instructional labs have been under considerable research and development. However, there seems to be no shared understanding of how the assessment of instructional labs should be arranged to best serve students' learning and development of expertise. This literature review intends to fill this gap by reviewing the research on classroom assessment of instructional labs from the perspectives of different objectives, purposes, and agents of assessment. The review reveals that classroom assessment in instructional labs has mostly focused on summative assessment, leaving the possibilities of formative assessment understudied. Further, assessment has been conducted mostly by teachers and teaching assistants, and the possibilities for students' participation in assessment remain unutilized. Two major gaps in the research on instructional labs were identified. The first gap concerns students' active participation in assessment. Given the active role that students have in the laboratory, their agency in assessment appears to be narrow. The second gap concerns the inclusion of metaskills and perspectives on lifelong learning and work life in assessment. We summarize our findings into a research-based model that assists in the consideration and balancing of different objectives, purposes, and agents in the design of classroom assessment of instructional labs.

DOI: 10.1103/PhysRevPhysEducRes.19.020601

I. INTRODUCTION

The development of physics instructional labs has attracted considerable attention in recent years [1]. Traditionally, the learning objectives of lab instruction have focused on learning physics content knowledge via, for example, providing hands-on experience with theoretical models learned in lectures [2]. Current research on instructional labs calls for a focus on learning objectives related to experimental physicists' actual work—that is, the development of experimental skills and the approaches and ways of thinking that experimentalists use [3,4]. Certain instructional methods have been highlighted to increase the alignment of instructions with the renewed learning objectives of labs. These include, for example, removing verification goals [2] and adding open-ended elements to lab tasks [5], with an increase in student decision-making to promote student agency [6,7].

The alignment of learning objectives and instruction is important, but we must also consider their alignment with the assessment [8]. As assessment is highly influential for learning [8,9], it is a key element to consider when developing all instruction. A scan of recent research reveals that assessment has not been at the center of the development of instructional labs. Individual researchers and research groups have conducted studies related to assessment practices (see especially Refs. [10–13]), but critical and ongoing discussions about the assessment of labs seem to be scant.

This article aims to draw attention to assessment as a key element in the development of instructional labs. We present results from a literature review aimed at determining the status quo of the research on the assessment of instructional labs. This study also reflects on the findings of the literature review by using general assessment literature as a comparison point. This reflection is aimed at identifying the strengths and gaps in the current research on assessment and instructional labs.

A. Creating a framework of assessment: Three perspectives

Assessment influences learning in many ways. Besides hindering or advancing learning, assessment demonstrates to students what is valued, what is worth learning, and even what *is* learning [8,9,14]. Assessment shapes students' learning approaches [15,16], affects how students use their limited time resources [17], and influences how students understand the nature of the subject in question [16]. Well-designed assessment can support students' self-efficacy [16], agency [18], and cooperation [19].

Here, we address the fact that the term "assessment" has two primary meanings. The first we call *classroom assessment*, which is the focus of this paper. By classroom assessment, we mean "assessments where the main decisions about what gets assessed, how the students will be assessed, and the scoring of the students' responses, is undertaken by those who are responsible for teaching the same students" [20]. Moreover, classroom assessment influences the same students who are assessed. If a student's performances appear as low, they receive constructive feedback and extra support—or fail the course.

Another meaning of assessment *is the measurement of learning outcomes for research and accountability purposes*, which, in the case of labs, is often done via assessment instruments, such as the Physics Lab Inventory of Critical Thinking (PLIC) [21] or Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [22]. The latter form of assessment is primarily used to provide information to instructors or researchers to develop their lab courses or conduct research. The instruments used are not designed to be used for grading or to provide feedback to students themselves. The aim of this type of assessment is not to influence the students who are being assessed; instead, future learning arrangements or assessments might be adjusted.

The practices of classroom assessment can be examined from various perspectives. In the following sections, we discuss three essential perspectives that form a literature-based framework for our literature review.

1. What is being assessed: The objectives of assessment

Researchers largely agree that three central elements of learning and teaching should be aligned: learning objectives, instruction, and assessment. This notation is called *constructive alignment* [8]. It signifies that we must consider the learning objectives that students practice, master, and demonstrate in a certain learning unit, and plan the instruction and assessment accordingly. When assessment and learning objectives are misaligned, students tend to focus on what is assessed rather than on what they are supposed to learn. If instruction and assessment are misaligned, students' efforts do not result in good grades, which lowers their motivation to participate.

What then are the learning objectives of instructional labs? The often cited recommendations for the undergraduate physics laboratory curriculum by the American Association of Physics Teachers (AAPT) [23] outline six focus areas, i.e., learning objectives for lab curricula: (i) Constructing knowledge, that is, the ability to collect, analyze, and interpret actual data and develop their physical worldview accordingly, (ii) modeling, (iii) designing experiments, (iv) developing technical and practical laboratory skills, (v) analyzing and visualizing data, and (vi) communicating physics. All these objectives can contribute to the development of students' experimental skills.

Learning physics content knowledge, including understanding and applying concepts, is a common learning objective of physics labs [2,4,24]. However, using labs to reinforce and confirm physics content presented in lectures is discouraged by research [4,24] as ineffective and even in conflict with the epistemology of physics. A more inductive approach to content learning is another matter, and using experimental observations to understand physical concepts and relationships between variables is in line with the epistemology of physics.

Metaskills are seen as an important objective of higher education [25], and in physics instructional labs, they play a central role [26]. Metaskills can be defined to include the categories of self-awareness, creative problem-solving, and resilience which are further divided into subcategories [27]. Inquiry skills are intertwined with metaskills [25]. For example, when designing experiments and developing technical and practical laboratory skills, students need awareness of their current skills and knowledge so that they can set their goals rationally. Besides metalevel skills, metaskills include also more general skills that are useful in life, such as confidence that belongs to the category of resilience [27].

2. What is assessment used for: Purposes of assessment

Classroom assessment is a combination of summative and formative assessment practices. *Summative assessment*, that is, assessment of learning, is the most traditional form of assessment [14]. It refers to a retrospective evaluation of what has been learned. In tertiary education, the most important purpose of summative assessment is the control and certification of learning. With summative assessment, institutions ensure that future professionals, for example, medical doctors, teachers, and physicists, can do their jobs, and for that purpose, pass versus fail is the crucial decision. However, besides controlling the pass or fail rate, most institutions also administer grades to their students. Measuring the level of learning is often emphasized in secondary education because heterogeneous student groups apply to various higher education institutions based on their secondary school grades. Given the high influence of summative assessment on students' life courses, summative assessment designs, especially their validity and reliability, have been widely researched and discussed (in physics, see Ref. [28]). There are also beliefs that grades motivate learning and function as influential feedback, but research has shown that grades do not work for those purposes [29].

Formative assessment, that is, assessment *for* learning, refers to an assessment that supports learning. Formative assessment can appear in many forms, such as feedback, prompts, questions, and dialogue. The form of assessment does not define whether the assessment is formative or summative but the purpose for which it is used [20]. For example, physics exams are formative if they are not only used to judge learning outcomes at the end of the course but are also used as part of normal coursework to provide information to both students and teachers on how students could be supported in their learning. Based on the results of formative assessment, teachers and teaching assistants (TAs) can adapt their teaching or students can adapt their ways of learning.

Besides summative and formative assessment, which consider the current needs of certifying and supporting learning, a third purpose of assessment has been raised in the literature: *sustainable assessment.* The concept was introduced by Boud [30] (and revisited by Boud and Soler [31]), and it goes beyond the purposes of immediate formative and summative purposes to include long-lasting (sustainable) outcomes of assessment and to serve the needs of lifelong learning and work life. For example, the course-specific aims of formative and summative assessment are nicely attained if the teacher or TA gives students detailed instructions, clear feedback, and justified grades. However, such an assessment does not prepare students for complex working lives, where answers and decisions are ambiguous, and workers must assess their own performance. Highly educated professionals need evaluative judgment, that is, the capability to evaluate their own work and understand requirements of high quality [32].

How can assessment be made sustainable? For example, developing students' ability to assess their own work, provide and ask for feedback and process it can be considered to serve the aims of sustainable assessment. These skills are needed in most working environments, and such skills also support students' lifelong learning. Two common ways to develop these skills are the use of exemplars and the use of self- and peer assessment. Researchers have argued that discipline-specific feedback practices are useful and they even suggest that processing conflicting and offensive feedback is profitable if that is what the feedback culture of the future profession will be [33]. The logic here is that learning to process feedback is so important for students' future lives and work, such that time should be invested to learn feedback processing, even when it might temporarily lower the efficiency of learning other goals.

3. Who is assessing: Agents of assessment

Assessment and feedback have long been considered the teachers' territory, but recent literature emphasizes that students should be key players in the assessment process [34]. One rationale for the change derives from constructive alignment. As students are expected to be active participants during learning, they should not be passive receivers in assessment [14]. One way to examine assessment is from the perspective of agents: Who is considered qualified to assess, who is supposed to deliver feedback, and whose judgment counts?

With formative assessment, the understanding of students' role has recently developed. Until the 2010s, the research used to emphasize the teacher's role and investigate what kind of feedback is most effective. Hattie and Timperley's meta-analysis of feedback research [35] is a seminal example of such an approach. They state, for example, that feedback should not focus on the person but on the task and learning process. Feedback should inform students about what they already know, what the learning goals are, and how those could be achieved, and that feedback about success is more effective than feedback about failure. Around 2010, the focus of research turned to students' role in formative assessment, and researchers began to examine how students receive, produce, process, and use feedback [36–39]. The main argument for the perspective change was that even perfect feedback does not benefit the receiver unless they process and use it. A new concept, *feedback literacy*, which refers to students' ability to benefit from feedback opportunities, was introduced [37,40] and was enthusiastically received by researchers. In this line of research, teachers' main responsibility is to create learning and assessment designs that allow students to generate and receive feedback and help them develop their abilities to benefit from such opportunities [41].

Students' participatory role in assessment has induced changes to teacher-led summative assessment, such as making learning goals and assessment criteria transparent [42,43]. Aside from teacher-led assessment, the use of self- and peer feedback in formative and summative assessment has been under substantial research. Typically, *self-assessment* is used in a formative way, meaning that students are guided to evaluate their work or skills in order to develop them, preferably with explicit assessment criteria. Self-assessment can also be used for summative purposes; thus, students decide on their own grades. Such a procedure has benefits, as it promotes a deep approach to learning and can improve students' learning outcomes [15]. There is also a considerable amount of research on summative peer assessment, which is often motivated by the intent to find research-based ways of entrusting time-consuming summative assessment to students. However, research suggests that formative peer assessment—that is, providing feedback to and receiving feedback from one's peers during learning—is more beneficial than summative peer assessment [44]. In peer assessment, both the assessor's and assessee's roles are beneficial for learning, but assessing peer's work and providing feedback for peers

TABLE I. Search criteria and matches for the databases.

Database	Search criteria	Matches
ERIC	("chemistry" OR "physics") AND ("experimental work" OR "practical work" OR "instructional labs" OR "lab course" OR "student lab" OR "student laboratory" OR "lab work") AND ("university" OR "higher education") AND ("assessment" OR "feedback" OR "grading" OR "graded" OR "grade" OR "grades")	64
WoS	("chemistry" OR "physics") AND ("experimental work" OR "practical work" OR "instructional labs" OR "lab course" OR "student lab" OR "student laboratory" OR "lab work") AND ("university" OR "higher education") AND ("assessment" OR "feedback" OR "grading" OR "graded" OR "grade" OR "grades")	97
Scopus	(TITLE-ABS-KEY (("chemistry" OR "physics") AND ("experimental work" OR "practical work" OR "instructional labs" OR "lab course" OR "student lab" OR "student laboratory" OR "lab work") AND ("university" OR "higher education") AND ("assessment" OR "feedback" OR "grading" OR "graded" OR "grade" OR "grades")) AND ABS ("assessment" OR "feedback" OR "grading" OR "graded" OR "grade" OR "grades"))	63
Physical Review Physics Education Research	("chemistry" OR "physics") AND ("experimental work" OR "practical work" OR "instructional labs" OR "lab course" OR "student lab" OR "student laboratory" OR "lab work") AND ("university" OR "higher education") AND ("assessment" OR "feedback" OR "grading" OR "graded" OR "grade" OR "grades")	4

seem to be even more beneficial for learning than receiving peer feedback [32].

Besides students and teachers, computers can take a role in summative and formative assessment, for example, by automatically assessing multiple-choice questions and providing grades or helping students to identify areas that require practice. Computers can be understood as agents of assessment [45,46], and we include them in our framework.

B. Research questions

The present review analyzes scientific articles that have investigated or carefully considered the assessment processes of instructional labs in higher education. Our review summarizes current knowledge, presents good practices, and identifies which areas of assessment need further research. The research aim is to provide guidance for people who design assessment of instructional labs by creating a model that assists in the design process. To achieve our aims, we address the following research questions:

RQ1: To what extent does the literature of instructional labs consider different objectives of assessment, that is, content, skills, and metaskills, and what are the main findings related to these?

RQ2: To what extent does the literature of instructional labs consider different purposes of assessment, that is, summative, formative, and sustainable assessment, and what are the main findings related to these?

RQ3: To what extent does the literature of instructional labs consider different agents of assessment, that is, teacher or TA, self, peer, and computer, and what are the main findings related to these?

II. METHODS

A. Literature search

We conducted a systematic literature search of articles related to classroom assessment of instructional labs. The selection of search terms required several test rounds and discussions until we found a combination that produced acceptably relevant results (Table I). Given that the number of studies concentrating on physics instructional labs was quite small, we decided to include studies from chemistry, which we consider to be close to physics in terms of the objectives and procedures of laboratory learning (see Ref. [47] for chemistry and Ref. [23] for physics). We limited the context of the studies to higher education and used various sets of terms to capture the studies of instructional labs and the topic of classroom assessment. We accepted only peer-reviewed articles or dissertations for the results. Ultimately, no dissertations fit our search criteria.

We conducted the literature review using three databases: ERIC, Web of Science (WoS), and Scopus. Additionally, we searched for articles published in Physical Review Physics Education Research directly on the journal website. The search criteria for ERIC, WoS, and Physical Review Physics Education Research were slightly different from those of Scopus, where the same research criteria would have produced 6220 matches that appeared largely irrelevant. Therefore, in Scopus, we focused the search so that a term relating to assessment ("assessment" or "feedback" or "grading" or "graded" or "grade" or "grades") was mandatory in the *abstract*, which improved the validity of the search results.

The final search was conducted on April 1, 2022, and it yielded 223 matches: ERIC—64, WoS—97, Scopus—63,

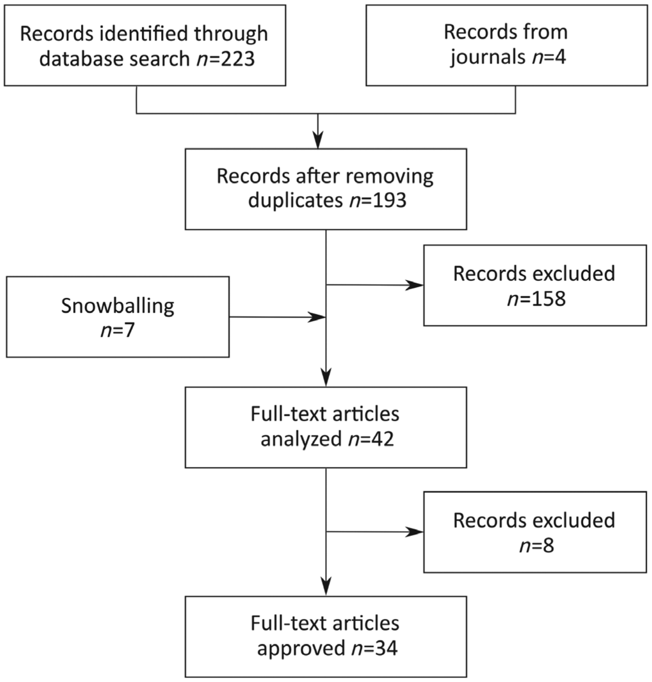

FIG. 1. Systematic search process.

and Physical Review Physics Education Research—4. After removing duplicates, we were left with 193 articles. We read through the articles and excluded those that did not match our scope. During this step, we were often able to exclude the paper based on the abstract, but sometimes the decision of exclusion versus inclusion required an examination of the full text. The reasons for exclusion were that the study was not an educational study ($n = 52$), the assessment was not mentioned or described in the study ($n = 37$), the subject was not physics or chemistry ($n = 21$), the context was not tertiary ($n = 19$), the study was not about instructional labs ($n = 15$), assessment referred to the measurement of learning results ($n = 8$), the article was not a research article ($n = 6$), we could not access the article ($n = 5$), and the article was not written in English ($n = 3$). After this exclusion round, we ended up with 35 articles.

We used snowballing to complement the search with relevant articles that we already knew about or identified from the other papers during the review. This added seven articles to the total. In total, we had 42 articles for the final data analysis. However, while reading the full articles, we still had to exclude eight articles, mainly because they did not sufficiently describe the assessment design for the needs of our analysis. In the end, 34 articles were selected for the literature review. Figure 1 visualizes our article selection process.

B. Data analysis

Each author read a third of the selected articles and analyzed them from the perspective of our research questions. Table II showcases the guide we created for the analysis. We collected the key findings of each article from a large summary table based on the analysis guide. During reading, we realized that some general information, that is, their publishing year and the rationale of the study might be needed to understand the articles better. Therefore, we added these two perspectives to the summary table.

TABLE II. Guide for analyzing articles.

Category	Subcategory	Instructions for reporting
Objectives of assessment	Content learning	If mentioned, describe which contents are assessed.
	Experimental skills	If mentioned, describe which skills are assessed.
	Metaskills	If mentioned, describe which meta-skills are assessed.
Purposes of assessment	Formative: Supporting students' learning	Describe the implementation of formative assessment.
	Summative: Measuring and verifying students' learning	Describe the implementation of summative assessment.
	Sustainable: Besides formative and summative purposes, the needs of working life, life, or lifelong learning have been mentioned.	Describe how the sustainable perspective was considered?
Agents of assessment	Teacher or teaching assistant	Specify, if the assessor is a teacher, TA, both, or unclear. Describe their role in assessment.
	Self	Describe their role in assessment.
	Peer	Describe their role in assessment.
	Computer	Describe their role in assessment.
General information	Publishing year	
	Rationale of the study	Describe, what has motivated the research or what is the gap that the study fills?

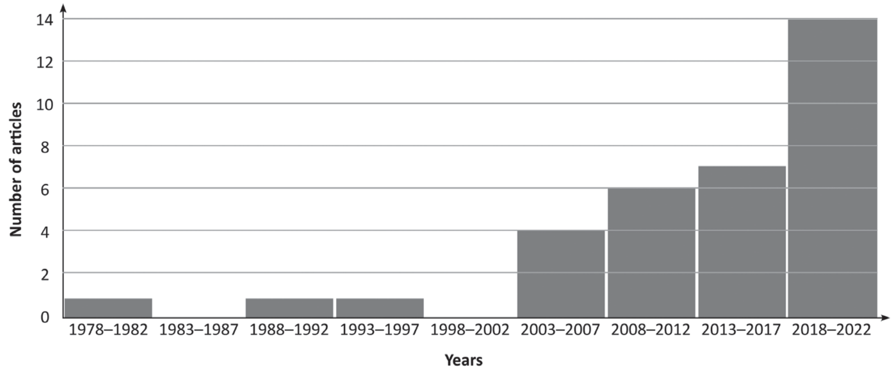

FIG. 2. Number of articles per year about the assessment of instructional labs.

III. FINDINGS

Based on the publishing years of the articles, the topic of classroom assessment in instructional labs has received increasing attention (Fig. 2). Despite the positive development, the topic is still marginal since a few researchers or research groups present a considerable amount of the whole research. This branch of research was started by Etkina *et al.* [10–13], and this research group, together with the research groups of Hensiek *et al.*, Towns *et al.* [48–50], and Pols *et al.* [51,52] presents more than a quarter of the literature and even a larger share of the studies that focus especially on assessment.

We found five different rationales for studying classroom assessment in instructional labs (Table III). In the first two, classroom assessment is not the main rationale. The first rationale is that some practical challenge(s) need to be resolved through research, and this research also touches upon classroom assessment. Such challenges include, for example, the lack of resources or changes to teaching caused by the COVID-19 pandemic. The second rationale is the need or aspiration to develop instructional labs; therefore, assessment is addressed. Topics range from developing a particular lab course or experimental method to developing the TA's role in assessment. The other three rationales explicitly focus on classroom assessment. The third rationale is to improve the constructive alignment of instructional labs, that is, the alignment between learning objectives, instruction, and assessment. The fourth rationale relates to challenges in the contemporary assessment of assessment practices, such as unequal grading practices or the high workload caused by current assessment practices. The fifth rationale is to improve learning in instructional labs in general by focusing on assessment practices. In addition to these five groups, two studies did not provide a clear rationale for their study.

A. RQ1: How does the literature consider different objectives of assessment?

Half of the studies mentioned experimental skills as an objective of assessment, as presented in Fig. 3. Content learning was mentioned as an objective in one-third of the

TABLE III. Rationales for considering assessment in studies of instructional labs.

Rationale of the study	Studies
Practical challenging situations drive the consideration of assessment	Brunauer (2016); Faulconer *et al.* (2018); Hoehn *et al.* (2021); Pols, (2020).
Assessment is considered during the design of instructional labs	Herrington and Nakhleh, (2003); Logan and Rumbaugh (2012); Matilainen *et al.* (2021); Meester and Maskill (1995); Roberts (1980); Read and Kable (2007); Gilani and Dushkina (2009); Sedumedi (2017); Hall (2014).
Assessment is considered for constructive alignment	Duis *et al.* (2013); Zwickl *et al.* (2012); El-Gabry (2021); Hensiek *et al.* (2017); Sena-Esteves *et al.* (2019); Luchembe and Shumba (2019).
Challenges of current assessment practices inspire considering new assessment	Avargil *et al.* (2019); Faletič and Planinšič (2020); Pols *et al.* (2022); Silva *et al.* (2009); Sobhanzadeh and Zizler (2021).
Assessment is considered to support learning	Etkina *et al.* (2009); Etkina *et al.* (2010); Etkina, Murthy, and Zou (2006); Etkina, Van Heuvelen *et al.* (2006); Hensiek *et al.* (2016); Veiga *et al.* (2019); Towns *et al.* (2015); Burkholder *et al.* (2021).
No clear rationale	Robyt and White (1990); Yang *et al.* (2021)

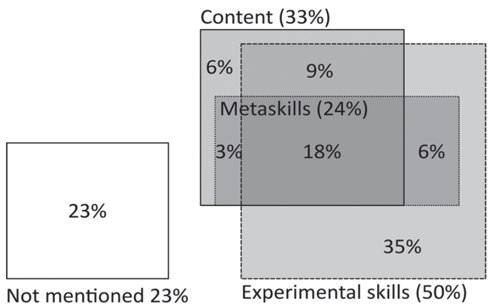

FIG. 3. Objectives of assessment in the articles. The size of a rectangle corresponds to the number of articles that consider the given objective. Overlapping rectangles refer to articles focusing on multiple objectives.

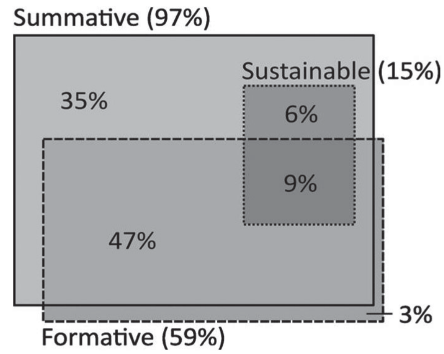

FIG. 4. Purposes of assessment in the articles. The size of a rectangle corresponds to the number of articles that consider the given purpose. Overlapping rectangles refer to articles focusing on multiple purposes.

studies, and meta-skills in one-fourth of the studies. Part of the studies considered all these aspects, and about one-fourth of the studies did not define the objectives of the assessment.

1. Content

Only 33% of the studies mentioned learning content as an objective of assessment [53–60]. The description of what "content" meant was rarely explicit, but a few studies from chemistry provided some details. In particular, Read and Kable's [57] article carefully describes the conceptual, experimental skill, and metalevel learning outcomes and the indicators that show the learning outcomes have been achieved. An example of their conceptual objective is, "Students will learn that chemical change can produce a change in temperature, and that, conversely, heating and cooling can induce chemical".

2. Experimental skills

Different experimental skills were the most mentioned objective of the assessment. Typical skills included designing an experiment [13,22,51,52,61,62], conducting experiments including technical lab skills [48,49,51–54,56,61,63–68], analyzing results and errors [10,12,13,51–54,57,60,62–66,68], and communicating results [12,13,54,57,61,62,65]. In chemistry, lab skills were often specific, such as electrophoretic separation [63] or using a volumetric flask [48,49]. Only studies from chemistry mentioned safety as an aspect of experimental skills [53,56,57,63,69]. Less common mentions were the use of imagination and the application of facts [68] and modeling [61].

3. Meta-skills

About a quarter of the studies included meta-skills as the objectives of the assessment. Teamwork [48,54,64] and negotiation and communication with peers [57] were the most common objectives. Other metaskills mentioned were cooperative learning [63], decision making [69], responsibility [53], general-level consideration of scientific principles [57], enduring mistakes [59], and general (not science related) presentation skills [66]. However, most of the studies did not specify how these aspects were considered in the assessment, but they mentioned them as general principles guiding the development of instructional labs.

B. RQ2: How does the literature consider different purposes of assessment?

About half of the articles considered both summative and formative purposes of assessment (Fig. 4). More than 40% did not consider formative assessment, while only 3% did not consider summative assessment. This implies that summative assessment was largely emphasized in the studies. Formative and summative assessment typically concentrate on the immediate needs of education—learning and validating course-specific skills—but besides these, we also examined how assessment considered the needs of work and lifelong learning. These perspectives were mentioned in five studies, but they were not explicitly articulated. The absence of work-life connections in assessment is surprising, considering that activities done in instructional labs are closely connected to many physicists' and chemists' future work.

1. Summative assessment

Studies reported various ways to implement summative assessment. The most common summative assessment task was a lab report [10,12,51–53,55,56,62,64,65,70–72], but in many studies, the lab report was only one assessment task, among others. Indeed, many assessment designs contained various assessment tasks [53,63,64,66,67,69,73]. Rubrics were used in several studies to assess lab reports [12,51,52,55,74]. The justification for the use of rubrics was to improve the reliability and transparency of the assessment. Summative prelab tests/assignments were used to ensure that students were prepared when they came to the lab [60,66,67,69]. Other types of summative assessment tasks were laboratory notebooks [53], worksheets [64], multiple-choice questions [11], posters [63,69], exams [54,66], practical exams [73,75], and laboratory data [58]. In some studies, students were able to use material, for example, their lab notes, on exams [54,68], and in other studies, the assessment task was conducted in a group [62,73].

Three articles [48–50] used videos of students' performance in labs as an assessment task. In their assessment design, summative and formative assessment were intertwined; the summative assessment was on a pass or fail scale, and students were able to redo the task until their performance was acceptable. With a failed performance, students received guiding feedback on how to improve their performance. Thus, the purpose of the assessment was summative if the performance was assessed as acceptable but formative if it was assessed as failed.

2. Formative assessment

The most common form of formative assessment was feedback from the teachers or TAs that was given on a draft or the final product, for example, a lab report [10,11,52,59,63,67,71,76]. In some studies, multiple iteration cycles were allowed and encouraged [59,74]. Peer assessment [13,63,69–71] and self-assessment [10–13] were used to support students' learning, often with rubrics [10–13,74]. Other forms of formative assessment were prompts [10,11] and class and group discussions [54,66]. Prompts were a kind of predesigned feedback in laboratory handouts that supported students in focusing on the essential aspects of experimental work. Discussions were used to encourage students to share their thoughts and questions and to recognize potential misconceptions.

3. Sustainable assessment

Few studies mentioned how assessment tasks could support the needs of work and lifelong learning. Burkholder *et al.* [69] stated that giving students the experience of experimentation decisions helps them conduct open-ended projects in the future. Gilani and Dushkina [66] recognized the importance of lifelong learning, and Sena-Esteves *et al.* [73] acknowledged the need to consider working life. However, the implementation of these aspects in assessment was not explicitly articulated in these studies. Etkina *et al.* [10] gave an explicit example of sustainable assessment: after each experiment, students were asked to think about how the experiment related to their future work by answering the question, "Why did we do this lab?" Another concrete example of linking assessment with working life was from Matilainen *et al.* [72]. In their study, laboratory course students received feedback on their seminar presentations from an outside chemistry specialist from the industry.

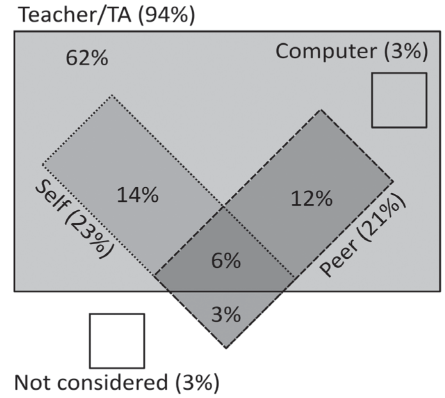

FIG. 5. Agents of assessment in the final articles. The size of a rectangle corresponds to the number of articles that consider the given agent. Overlapping rectangles refer to articles focusing on multiple agents.

C. RQ3: How does literature consider different agents of assessment?

As shown in Fig. 5,[1] the most common agents of assessment were teachers and/or TAs. Only 3% of the studies, that is, one study [13], focused only on students as assessors, while 62% of the studies focused only on teachers or TAs. In one-third of the studies, students were considered agents in assessment, which meant that they assessed themselves (23%) and/or their peers (21%).

1. Teachers and teaching assistants

Overwhelmingly, in the reviewed studies, the most common agent of assessment was a teacher or TA, who summatively assessed students' lab reports, exams, or other products, or made the course evaluation. It was also common in the studies that teachers or TAs provided formative feedback. In several studies, the exact assessment procedure was not clearly described, but the teacher or TA was implied as the assessment agent.

2. Peers

Peers' most common role in assessment was to provide formative feedback [13,54,69,70]. In one study [13], students first used rubrics in peer assessment to practice their evaluative judgment and later used them in self-assessment. In another study [69], students answered questions about their group members and met face-to-face with them, providing feedback with the help of structured statements. Only in two studies did students play an active role in summative assessment. In Brunauer's [63] and Burkholder *et al.*'s [69] study, students assessed each other's participation in teamwork, and these evaluations had a small influence on their final grades. Duis *et al.* [53]

[1]Although not directly referenced in the text, the following articles [77,78] were included in the review study, including Figs. 3, 4, and 5.

mentioned "peer marking," which seems to be a summative procedure, but the practice was not explained.

3. Self

Most often, self-assessment was used to support students' learning during the experimental process and to guide them in producing feedback on how to improve their work. Rubrics were a common tool in self-assessment [10–13,74]. In one study, self-assessment was used at the end of the course [66]. In another study [59], students participated in assessment by choosing quizzes/activities that would influence the final summative grade. One study [63] instructed students to ask for feedback from a TA, which gave them agency over their own formative assessment. They did not give feedback themselves but recognized the moments in which they needed feedback about their work.

4. Computers

In Silva *et al.*'s [58] study, a computer was an agent of assessment. In this study, students submitted the results of their lab assignment (determining the concentration of a certain chemical in a solution) to the assessment system, and the students' input was automatically graded by the system. As the study was published over a decade ago, we found it surprising that despite several advantages reported and the development of digital learning environments, there were no other studies using computers as assessors in instructional labs.

IV. DISCUSSION

Based on our review, classroom assessment of instructional labs is summatively emphasized, instructor-led, and focused on experimental skills and physics content. The scenery can be described with the word "traditional." However, the reviewed articles presented several new and experimental approaches to assessment, which is natural, since research tends to focus on new designs rather than traditional ones. Despite the new features that the assessment designs presented, they did not widely challenge or question the traditional presumptions of assessment.

In the reviewed articles, we identified one main renewal: as experimental skills have gained ground as learning objectives in instructional labs, at least in North America [1], research on assessment has also begun to understand experimental skills as the principal objective of the assessment. However, assessing skills is more complex than assessing knowledge; therefore, the assessment designs of instructional labs need to be restructured, and this process is still in the beginning. Further, a common understanding of *how* those skills should be assessed is lacking. Another gap in the research is the inclusion of metaskills in the assessment. If such skills are included, they add one more layer of complexity to the assessment.

What has not changed in instructional labs' assessment is the emphasis on summative assessment, and teachers' or TA's dominating role. Teachers or TAs are responsible for summative assessments (pass or fail and grades), and they also spend a considerable amount of time providing feedback. If more complex skills are to be included in the assessment, we must consider who could best assess them? Can the instructor provide feedback about the student group's experimental process or collaboration or grade these skills reliably? Not necessarily.

High reliability is better achieved with simple skills. Memorization and repetition of knowledge can be assessed with great reliability; however, they are not sufficient skills for highly educated experts. Creative application of knowledge and skills and evaluation of one's own performance are higher-order skills that should be targeted, but they are hard to assess reliably. Assessment requires compromises [79]. If we want to improve the validity of the assessment, we need to factor in complex objectives that matter in physicists' work lives, which necessarily means loosening up the reliability of grading.

How can the dilemma be solved? First, the conflict does not concern formative assessment and feedback, since feedback does not need to be comparable. We encourage designers of assessment to change the emphasis from summative assessment taking place in the course to feedback processes and the examination of learning during learning. Second, we recommend giving students agency in assessment in general. Contemporary research on assessment provides justified solutions for such arrangements. In the following section, we consider some of these.

A. The model for designing assessment

The principal aim of our review was to provide guidance for creating an assessment design for instructional labs. After conducting the review, we can conclude that the literature is too narrow to create a "how-to-assess-labs" model, despite several important insights and ideas. Moreover, the articles did not use research designs that would have allowed making preferences between different assessment practices. By combining the general assessment literature with that of instructional labs, we propose a model in the form of a flowchart that offers recommendable aspects to consider when designing the assessment of instructional labs. These are presented in Fig. 6 and explained in the following sections.

B. Components of the assessment model

1. Objectives

In our model, the process of designing an assessment begins with the choice of assessment objectives. The current understanding of instructional labs is that they are more efficient for learning experimental skills than learning content knowledge, and based on our review, the

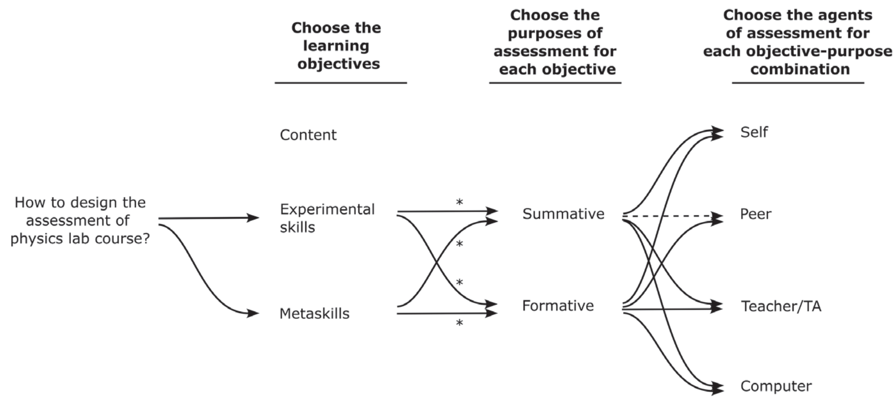

FIG. 6. Flowchart that presents the combinations of objectives, purposes, and agents to consider when designing the assessment of instructional labs. The perspective of sustainable assessment is embedded in other processes and marked with *. The dashed lines present a combination that, based on literature, is usable but less recommendable for learning.

research on assessment reflects this. Researchers have already suggested sets of skills that could be utilized when choosing objectives and designing criteria for their assessment [10,13,51,65]. However, the research on the objectives of experimental skills in instructional labs is by no means saturated, and we call for new conceptualizations of skills and criteria that can assist teachers and students in designing, guiding, and assessing learning.

Instructional labs require and provide opportunities for practicing meta-skills, because learning is often cooperative and tasks open-ended, therefore asking for self- and coregulation and creativity. We encourage teachers and researchers to consider including metaskills in assessment, even though their inclusion in instructional labs' assessment has been scant thus far. This recommendation derives from the alignment of instruction and assessment. If cooperation, for example, is valuable for learning, it is rational to consider it in assessment [19]. The inclusion of metaskills adds sustainability to assessment, as it guides students to reach beyond the immediate task at hand. Examples of metaskills in our review were El-Gabry's [64], "I know how to work as a team" and Read and Kable's [57] "One-on-one communication, explanation, and negotiation skills with a peer."

2. Purposes

The main purposes of assessment are formative (assessment for learning) and summative (assessment of learning). When combined with objectives, the requirement for formative assessment is that students receive feedback about each learning objective during learning and that the summative assessment consists of consideration of all or certain predetermined subsets of learning objectives. Therefore, in the purpose phase (Fig. 6), the principal question is which objectives should be chosen as the basis for summative assessment.

The third purpose, sustainable assessment, refers to serving aspects of lifelong learning and work life. In Fig. 6, we chose not to write sustainable assessment as its own category, but instead, we marked the elements of assessment in which these aspects could be supported. One possibility is choosing meta-skills as objectives of the assessment and assessing them formatively and summatively. Further, formative assessment, particularly feedback on experimental skills, can be arranged in ways that support students' work-related skills and feedback skills. We will return to these possibilities in the next section.

3. Agents

Instructional labs demand students' active participation, sometimes individually but more often in groups. With instructional labs, there has been an endeavor to move from cookbook experiments, in which students follow instructions, to more open-ended and creative research settings [49,68]. We draw on the idea of constructive alignment [8] to argue that because learning requires students' active participation, they should have agency in assessment as well. Two main roles for students in assessment are assessing themselves (self-assessment) and others (peer assessment). Designers of assessment should weigh the kind of participation in the assessment that best serves the students. Another obvious group of agents of assessment is teachers and/or TAs. They must participate in assessment, at minimum, by organizing assessment and guiding students to do it. In the phase of agents, the decisions are who (all) participate in the summative assessment of chosen objectives, and who (all) participate in the provision of feedback.

C. Combinations of purposes and agents

Examination of all combinations of purposes and agents shows that research on instructional labs' assessment has certain emphases and various gaps. In the following sections, we summarize the findings concerning all combinations and discuss the gaps by referring to what is known based on general assessment literature.

1. Teachers or TAs

Summative assessment.—Based on our review, the most common summative assessors in instructional labs are teachers or TAs. Even though teachers' or TAs' role in summative assessment appears to be the norm, it is not without challenges. Assessment between teachers or TAs is not necessarily consistent [55,58], the time they spend on grading needs consideration [12,58,65], and students' absence from labs calls for new solutions [59]. Despite these challenges, teachers' or TAs' participation in summative assessment is justified as they understand the quality of good performance, and they are responsible for the certification of students' learning. However, TAs' assessment skills must be ensured by sufficient training to ensure the quality of both summative and formative assessments [80]. Referring to Villarroel *et al.* [81], we recommend that to improve the transparency and reliability of assessment, summative assessment should always be based on assessment criteria and those would be introduced to students, for example, at the beginning of the lab course.

Formative assessment.—Teachers' or TAs' participation is common in formative assessment, too. In the articles we reviewed, teachers or TAs held feedback dialogues, provided informal feedback during lab work, and commented on students' drafts during the learning process. What is useful to remember when choosing where to invest resources is that delivering feedback alongside a grade appears largely inefficient [82]. Feedback is most productive during learning when students can still use it to improve their performance [41]. What was largely lacking in the articles we reviewed was organized support for students' feedback skills and processes. Students should not just be ordered to perform self- and peer assessment, but those skills should be practiced. We recommend that teachers or TAs invest more in meta-level discussions about feedback with students: Why is feedback delivered? Why are self- and peer assessment used? What does good feedback look like? How does one evaluate and use feedback? Such discussions can be more influential for learning than teachers' or TAs' investments in feedback comments. For planning, we suggest the work of Ketonen [83], which introduced a training program in the context of secondary school. In addition to self- and peer assessment, students' evaluative judgment and feedback skills can be promoted by showing and discussing exemplars of different levels of performance and by using rubrics that model the elements of quality [32].

2. Students

Summative assessment.—In our review, no study examined the possibility of letting students assess their work summatively. However, letting students decide on their own grades is justified and even recommendable if they receive proper training and support. Based on research, summative self-assessment, also known as self-grading, can promote students' self-efficacy and a deep approach to learning [15]. However, the practice is not without concerns, because influencing one's own grade, even by a small percentage, can cause students to give themselves unrealistically high grades [84]. This would be problematic from the perspective of accountability. The consequences of self-grading are intertwined with contextual factors [85], for example, whether the assessment culture is high stakes or low stakes [86]. We do not introduce summative self-assessment as a method that could be immediately implemented in all circumstances, but rather as a long-term direction to consider. Such an objective may require development of the local or national assessment practices and policies.

The dynamics of peer assessment is different from that of self-assessment. Even though the reliability of peer assessment with appropriate arrangements can be as high as teachers' assessment [44], it seems that summativity in peer assessment has the opposite effect on learning as it does in self-assessment. Whereas in self-assessment, the responsibility of an individual's own grade makes them examine their learning more intensively, in peer assessment, the orientation tends to turn from assisting peer's learning to more superficial deliverance of grades [44]. In addition, summative peer assessment does not increase one's ownership of one's own learning. It is possible that the future will bring new innovative approaches to summative peer assessment, but based on current research, we recommend that if students are to participate in summative assessment, it would rather be through summative self-assessment than summative peer assessment. However, peer assessment is a justified solution if it is primarily used to reduce teachers' or TAs' workload.

If metaskills, such as cooperation, self-regulation, and communication, are chosen to be assessed summatively, students should be allowed to assess them, because they have often followed the process from a closer distance than teachers or TAs. This could mean individuals assessing their own or their peers' work, or groups assessing their collective performance.

Formative assessment.—In formative assessment, as opposed to summative assessment, the articles included in the review placed students more often as active participants. Students assessed their own work, provided feedback about peers' drafts, and had feedback dialogues. Peer and self-assessment are essential components of formative assessment [87]. In these activities, students learn by using assessment rubrics, by evaluating their own and others' work, and by receiving feedback about their work. It is essential to remember that both self- and peer-assessment skills need practice [88] and practice requires time. We argue that the facilitation of feedback processes is perhaps the most important task of teachers or TAs in formative assessment, and we will discuss this in the next section. The rationale is that with such support, students' ability to use feedback improves (which also concerns feedback provided by teachers or TAs), students' ability to provide quality feedback improves, and their need for external feedback decreases. This kind of formative assessment is also sustainable because all the above-mentioned skills equip students for postuniversity life.

3. Computers

The use of computers as assessors will most likely increase in the upcoming years. In summative assessment, computer-based assessment reduces teachers' workload [46,89], and after getting used to computer-based assessment, students prefer it to pen-and-paper assessment [46]. In physics labs, students mainly practice higher-level skills that computers still lack the ability to assess [46], but the situation may change with the rapid development of AI.

The possibilities of computer-based assessment are even wider than those of summative assessment. Computers can offer automated feedback and adapt the assessment to students' responses, which can improve both formative and summative assessment [45,89,90]. Considering the growing possibilities of computers, we see that they will increasingly be used as assessors, also in instructional labs.

4. Sustainable perspectives

Formative assessment holds more possibilities for sustainable assessment than summative assessment does, because in postuniversity life, there will be no grades and little summative assessment, but feedback remains. One possibility of assessment to support future life is authentic feedback, which has features that are typical of the future profession [33]. With instructional labs, this could mean peer reviewing each other's lab reports and commenting on presentations, and having professional-like feedback dialogues with the working group. One possibility of authentic feedback, as suggested by Matilainen *et al.* [72], is receiving feedback from an authentic professional working in the industry.

The possibilities of creating connections to work life in instructional labs are wide. Study material can guide students to make links between the assessment task and their future profession [10], the assessment tasks can relate to, or even solve, authentic problems, or students could share their results in some form of media or participate in professional or public discussion. One way to make assessment more sustainable is to design assessment tasks that allow students to reach outside the classroom in some way [91]. Holmes and Wieman [92] suggest that placing emphasis on the quality of students' processes in the lab rather than the products prepares students better for research internships. The work of experimental physicists is closer to what is happening in instructional labs than what is happening in lectures. Therefore, instructional labs provide excellent possibilities for sustainable assessment, and we hope that more research will examine those in the future.

D. Questions for designing assessment

We have created questions that can be followed when creating an assessment design for instructional labs. We suggest that designers follow the steps in numerical order but return to the previous step if a need to adjust the previous choices appears. Designing an assessment means balancing different factors, such as policy, pedagogy, resources, validity, and reliability, and finding a satisfactory compromise may need several rounds of iterations.

1. Choice of objectives

- What are the learning objectives of the course?
  - Are some metaskills included?
- Which of these objectives (if not all) is used as a basis for summative assessment?

2. Planning summative assessment

- What are the pass or fail criteria?
- Who assesses whether the performance is acceptable or fails?
  - If teachers or TAs, how are they trained to do it?
  - If students, how are they trained to do it?
  - If students, how are they supported during assessment?
- If grades are used, what are the criteria for different grades?
- Who decides a student's grade?
  - If teachers or TAs, how are they trained to do it?
  - If students, how are they trained to do it?
  - If students, how are they supported during assessment?

3. Planning formative assessment

- What is the optimal timing of feedback?
  - When would formative assessment best support students' learning?
- Who provides feedback?
  - Self-assessment
    - How are students trained to do this?
    - How are students supported during assessment?
    - What kinds of rubrics are used?
  - Peer assessment
    - How are students trained to do this?
    - How are students supported during assessment?
    - What kinds of rubrics are used?
    - How are students guided to receive feedback?
  - Teacher or TA
    - What objectives does teacher's or TA's feedback focus on?
    - How are students guided to receive feedback?
  - Are there any other feedback agents?
    - Software, external professionals, etc.

V. LIMITATIONS

Our research has certain limitations. First, in database research, we used tighter search criteria for Scopus because of the expansive number of article matches. Even though the matches were largely irrelevant, it is possible that we missed some useful articles. In addition, the number of duplicates between the databases was relatively small, meaning that we probably did not identify all the relevant articles.

Another limitation is that many articles that matched the final criteria provided little detail about assessment arrangements. This is a common observation in reviews; for example, Topping [93] regretted inadequate details about peer assessment and offered a checklist about aspects of what to report. We did not provide such a list unless our reviewing guide (Table II) is considered as such. The lack of details was partly explained by the fact that assessment was not the focus of all reviewed articles. Vague descriptions left room for various interpretations, but we solved unclear cases through a joint discussion.

VI. CONCLUSION

Changing assessment practices is hard because they are intertwined with our disciplinary traditions and with our conceptions of learning. Moreover, previous experiences with assessment have built our understanding of the nature of learning, which is one reason why traditional ways of assessing appear coherent, whereas new solutions are hard to accept. Not only are teachers reluctant to change but students also tend to be accustomed to traditional assessment practices and, for example, reject the active role they are given [44]. However, as assessment guides learning powerfully, aligning assessment designs with objectives and instruction is worthwhile. Based on our review, there is a need for further research on how to implement assessment of instructional labs. We call for new innovative assessment designs that build on contemporary research and boldly challenge the traditional approaches. We hope that the model presented in this article provides tools for examining current practices and developing new ones.

[1] N. G. Holmes and H. J. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, Phys. Rev. Phys. Educ. Res. **16**, 020162 (2020).

[2] E. Smith and N. Holmes, Best practice for instructional labs, Nat. Phys. **17**, 662 (2021).

[3] C. Walsh, H. J. Lewandowski, and N. G. Holmes, Skills-focused lab instruction improves critical thinking skills and experimentation views for all students, Phys. Rev. Phys. Educ. Res. **18**, 010128 (2022).

[4] N. G. Holmes, J. Olsen, J. L. Thomas, and C. E. Wieman, Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content, Phys. Rev. Phys. Educ. Res. **13**, 010129 (2017).

[5] B. R. Wilcox and H. J. Lewandowski, Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics, Phys. Rev. Phys. Educ. Res. **12**, 020132 (2016).

[6] N. G. Holmes, B. Keep, and C. E. Wieman, Developing scientific decision making by structuring and supporting student agency, Phys. Rev. Phys. Educ. Res. **16**, 010109 (2020).

[7] Z. Y. Kalender, E. Stump, K. Hubenig, and N. G. Holmes, Restructuring physics labs to cultivate sense of student agency, Phys. Rev. Phys. Educ. Res. **17**, 020128 (2021).

[8] J. B. Biggs and C. S. Tang, *Teaching for Quality Learning at University: What the Student Does*, 4th ed. (McGraw-Hill, Society for Research into Higher Education & Open University Press, Maidenhead, England, New York, NY, 2011).

[9] B. Wisniewski, K. Zierer, and J. Hattie, The power of feedback revisited: A meta-analysis of educational feedback research, Front. Psychol. **10**, 3087 (2020).

[10] E. Etkina, A. Karelina, S. Murthy, and M. Ruibal-Villasenor, Using action research to improve learning and formative assessment to conduct research, Phys. Rev. ST Phys. Educ. Res. **5**, 010109 (2009).

[11] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, J. Learn. Sci. **19**, 54 (2010).

[12] E. Etkina, S. Murthy, and X. Zou, Using introductory labs to engage students in experimental design, Am. J. Phys. **74**, 979 (2006).

[13] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[14] C. Gipps, Socio-cultural aspects of assessment, Rev. Res. Educ. **24**, 355 (1999).

[15] J. H. Nieminen, H. Asikainen, and J. Rämö, Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models, Stud. Higher Educ. **46**, 1296 (2021).

[16] E. Panadero and A. Jonsson, The use of scoring rubrics for formative assessment purposes revisited: A review, Educ. Res. Rev. **9**, 129 (2013).

[17] K. Silseth and Ø. Gilje, Multimodal composition and assessment: A sociocultural perspective, Assess. Educ. **26**, 26 (2019).

[18] J. H. Nieminen and L. Tuohilampi, 'Finally studying for myself'—Examining student agency in summative and formative self-assessment models, Assess. Eval. High. Educ. **45**, 1031 (2020).

[19] D. Boud and M. Bearman, The assessment challenge of social and collaborative learning in higher education, Educ. Philos. Theory **0**, 1 (2022).
[20] P. Black and D. Wiliam, Classroom assessment and pedagogy, Assess. Educ. **25**, 551 (2018).
[21] C. Walsh, K. N. Quinn, C. Wieman, and N. G. Holmes, Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking, Phys. Rev. Phys. Educ. Res. **15**, 010135 (2019).
[22] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, Development and validation of the Colorado learning attitudes about science survey for experimental physics, AIP Conf. Proc. **1513**, 442 (2013).
[23] J. Kozminski *et al.*, AAPT recommendations for the undergraduate physics laboratory curriculum, Am. Assoc. Phys. Teach. **29**, 2 (2014).
[24] C. Wieman and N. G. Holmes, Measuring the impact of an instructional laboratory on the learning of introductory physics, Am. J. Phys. **83**, 972 (2015).
[25] H. Muukkonen and M. Lakkala, Exploring metaskills of knowledge-creating inquiry in higher education, Int. J. Comput.-Support. Collab. Learn. **4**, 187 (2009).
[26] K. L. Van De Bogart, D. R. Dounas-Frazer, H. J. Lewandowski, and M. R. Stetzer, Investigating the role of socially mediated metacognition during collaborative troubleshooting of electric circuits, Phys. Rev. Phys. Educ. Res. **13**, 020116 (2017).
[27] P. Prasittichok and K. K. Klaykaew, Meta-skills development needs assessment among undergraduate students, Heliyon **8**, e08787 (2022).
[28] M. Scott, T. Stelzer, and G. Gladding, Evaluating multiple-choice exams in large introductory physics courses, Phys. Rev. ST Phys. Educ. Res. **2**, 020102 (2006).
[29] A. A. Lipnevich and J. K. Smith, Response to assessment feedback: The effects of grades, praise, and source of information, ETS Res. Rep. Ser. **2008**, 1 (2008).
[30] D. Boud, Sustainable Assessment: Rethinking assessment for the learning society, Stud. Contin. Educ. **22**, 151 (2000).
[31] D. Boud and R. Soler, Sustainable assessment revisited, Assess. Eval. High. Educ. **41**, 400 (2016).
[32] J. Tai, R. Ajjawi, D. Boud, P. Dawson, and E. Panadero, Developing evaluative judgement: Enabling students to make decisions about the quality of work, High. Educ. **76**, 467 (2018).
[33] P. Dawson, D. Carless, and P. P. W. Lee, Authentic feedback: Supporting learners to engage in disciplinary feedback practices, Assess. Eval. High. Educ. **46**, 286 (2021).
[34] D. Boud, Shifting views of assessment: From secret teachers' business to sustaining learning, *Advances and Innovations in University Assessment and Feedback* (Edinburgh, 2014; online edn, Edinburgh Scholarship Online, 2014), p. 13.
[35] J. Hattie and H. Timperley, The power of feedback, Rev. Educ. Res. **77**, 81 (2007).
[36] D. Boud and E. Molloy, Rethinking models of feedback for learning: The challenge of design, Assess. Eval. High. Educ. **38**, 698 (2013).
[37] D. Carless and D. Boud, The development of student feedback literacy: Enabling uptake of feedback, Assess. Eval. High. Educ. **43**, 1315 (2018).
[38] A. Jonsson, Facilitating productive use of feedback in higher education, Act. Learn. High. Educ. **14**, 63 (2013).
[39] N. E. Winstone, R. A. Nash, J. Rowntree, and M. Parker, 'It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience, Stud. Higher Educ. **42**, 2026 (2017).
[40] P. Sutton, Conceptualizing feedback literacy: Knowing, being, and acting, Innov. Educ. Teach. Int. **49**, 31 (2012).
[41] D. Carless and N. Winstone, Teacher feedback literacy and its interplay with student feedback literacy, Teach. High. Educ. **28**, 150 (2023).
[42] G. Rodríguez-Gómez and M. S. Ibarra-Sáiz, Assessment as learning and empowerment: Towards sustainable learning in higher education, in *Sustainable Learning in Higher Education: Developing Competencies for the Global Marketplace*, edited by M. Peris-Ortiz and J. M. Merigó Lindahl (Springer International Publishing, Cham, 2015), pp. 1–20.
[43] *Assessment as Learning: Maximising Opportunities for Student Learning and Achievement*, edited by Z. Yan and L. Yang (Routledge, London, 2021).
[44] E. Panadero, Is It Safe? Social, Interpersonal, and Human Effects of Peer Assessment: A Review and Future Directions, in *Handbook of Social and Human Conditions in Assessment* (Routledge, New York, 2016), pp. 247–266.
[45] U. Mertens, B. Finn, and M. A. Lindner, Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis, J. Educ. Psychol. **114**, 1743 (2022).
[46] K. Perry, K. Meissel, and M. F. Hill, Rebooting assessment. Exploring the challenges and benefits of shifting from pen-and-paper to computer in summative assessment, Educ. Res. Rev. **36**, 100451 (2022).
[47] N. Reid and I. Shah, The role of laboratory work in university chemistry, Chem. Educ. Res. Pract. **8**, 172 (2007).
[48] S. Hensiek, B. K. DeKorver, C. J. Harwood, J. Fish, K. O'Shea, and M. Towns, Improving and assessing student hands-on laboratory skills through digital badging, J. Chem. Educ. **93**, 1847 (2016).
[49] S. Hensiek, B. K. DeKorver, C. J. Harwood, J. Fish, K. O'Shea, and M. Towns, Digital badges in science: A novel approach to the assessment of student learning, J. Coll. Sci. Teach. **46**, 28 (2017).
[50] M. Towns, C. J. Harwood, M. B. Robertshaw, J. Fish, and K. O'Shea, The digital pipetting badge: A method to improve student hands-on laboratory skills, J. Chem. Educ. **92**, 2038 (2015).
[51] C. F. J. Pols, P. J. J. M. Dekkers, and M. J. de Vries, Defining and assessing understandings of evidence with the assessment rubric for physics inquiry: Towards integration of argumentation and inquiry, Phys. Rev. Phys. Educ. Res. **18**, 010111 (2022).
[52] F. Pols, A physics lab course in times of COVID-19, Electron. J. Res. Sci. Math. Educ. **24**, 172 (2020).

[53] J. M. Duis, L. L. Schafer, S. Nussbaum, and J. J. Stewart, A process for developing introductory science laboratory learning goals to enhance student learning and instructional alignment, J. Chem. Educ. **90**, 1144 (2013).

[54] M. L. Hall and D. Vardar-Ulu, An inquiry-based biochemistry laboratory structure emphasizing competency in the scientific process: A guided approach with an electronic notebook format, Biochem. Mol. Biol. Educ. **42**, 58 (2014).

[55] S. Avargil, M. Bruce, S. Klemmer, and A. Bruce, A professional development activity to help teaching assistants work as a team to assess lab reports in a general chemistry course, Isr. J. Chem. **59**, 536 (2019).

[56] J. L. Logan and C. E. Rumbaugh, The chemistry of perfume: A laboratory course for nonscience majors, J. Chem. Educ. **89**, 613 (2012).

[57] J. R. Read and S. H. Kable, Educational analysis of the first year chemistry experiment thermodynamics think-in': An ACELL Experiment, Chem. Educ. Res. Pract. **8**, 255 (2007).

[58] A. Silva, R. Gonzales, and D. P. Brennan, Online grading of calculations in general chemistry laboratory write-ups, J. Educ. Technol. Syst. **38**, 155 (2009).

[59] M. Sobhanzadeh and P. Zizler, Selective assessment in introductory physics labatorials, Phys. Teach. **59**, 114 (2021).

[60] N. Veiga, F. Luzardo, K. Irving, M. N. Rodriguez-Ayan, and J. Torres, Online pre-laboratory tools for first-year undergraduate chemistry course in uruguay: student preferences and implications on student performance, Chem. Educ. Res. Pract. **20**, 229 (2019).

[61] B. Zwickl, N. Finkelstein, and H. J. Lewandowski, Transforming the advanced lab: Part I—Learning goals, AIP Conf. Proc. **1413**, 391 (2012).

[62] Y. Yang, Y. Zhang, X. Xiong, W. Zhang, W. Chen, and S. Ge, From lab scale to mass production: A project-based learning on the preparation of (S)-epichlorohydrin for enhancing college student engineering practical abilities, J. Chem. Educ. **98**, 3804 (2021).

[63] L. Brunauer, Purification and electrophoretic characterization of lactate dehydrogenase from mammalian blood: A different twist on a classic experiment, J. Chem. Educ. **93**, 1108 (2016).

[64] L. El-Gabry, Development, implementation and assessment of thermodynamics lab kits for remote lab instruction in *ASEE Annual Conference and Exposition, Conference Proceedings* (2021).

[65] S. Faletič and G. Planinšič, How the introduction of self-assessment rubrics helped students and teachers in a project laboratory course, Phys. Rev. Phys. Educ. Res. **16**, 020136 (2020).

[66] T. H. Gilani and N. M. Dushkina, The Undergraduate Optics Course at Millersville University, Education and Training in Optics and Photonics, OSA Technical Digest Series (2009), https://doi.org/10.1364/ETOP.2009.ETA4.

[67] M. A. M. Meester and R. Maskill, First-year chemistry practicals at universities in England and Wales: Organizational and teaching aspects, Int. J. Sci. Educ. **17**, 705 (1995).

[68] J. F. Robyt and B. J. White, Laboratory practical exams in the biochemistry lab course, J. Chem. Educ. **67**, 600 (1990).

[69] E. Burkholder, L. Hwang, E. Sattely, and N. Holmes, Supporting decision-making in upper-level chemical engineering laboratories, Educ. Chem. Eng. **35**, 69 (2021).

[70] J. R. Hoehn, M. F. J. Fox, A. Werth, V. Borish, and H. J. Lewandowski, Remote advanced lab course: A case study analysis of open-ended projects, Phys. Rev. Phys. Educ. Res. **17**, 020111 (2021).

[71] D. Luchembe and O. Shumba, Practical work, simulations and feedback to address undergraduate physics students' challenges in understanding circular and rotational motion, Educ. Q. Rev. **2**, 585 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3450027.

[72] R. Matilainen, P. Nuora, and P. Valto, Student experiences of project-based learning in an analytical chemistry laboratory course in higher education, Chem. Teach. Int. **3**, 229 (2021).

[73] T. Sena-Esteves, C. Morais, A. Guedes, I. B. Pereira, M. M. Ribeiro, F. Soares, and C. P. Leão, Students' perceptions regarding assessment changes in a fluid mechanics course, Educ. Sci. **9**, 152 (2019).

[74] S. Faletič and G. Planinšič, How the introduction of self-assessment rubrics helped students and teachers in a project laboratory course, Phys. Rev. Phys. Educ. Res. **16**, 020136 (2020).

[75] E. K. Faulconer, J. C. Griffith, B. L. Wood, S. Acharyya, and D. L. Roberts, A comparison of online and traditional chemistry lecture and lab, Chem. Educ. Res. Pract. **19**, 392 (2018).

[76] A. R. Roberts, Foundation physics at Liverpool Polytechnic, Phys. Educ. **15**, 191 (1980).

[77] D. G. Herrington and M. B. Nakhleh, What defines effective chemistry laboratory instruction? Teaching assistant and student perspectives, J. Chem. Educ. **80**, 1197 (2003).

[78] T. D. T. Sedumedi, Practical work activities as a method of assessing learning in chemistry teaching, Eurasia J. Math. Sci. Technol. Educ. **13**, 1765 (2017).

[79] Y. Xu and G. T. L. Brown, Teacher assessment literacy in practice: A reconceptualization, Teach. Teach. Educ. **58**, 149 (2016).

[80] N. Wald and T. Harland, Rethinking the teaching roles and assessment responsibilities of student teaching assistants, J. Furth. High. Educ. **44**, 43 (2020).

[81] V. Villarroel, S. Bloxham, D. Bruna, C. Bruna, and C. Herrera-Seda, Authentic assessment: creating a blueprint for course design, Assess. Eval. High. Educ. **43**, 840 (2018).

[82] A. A. Lipnevich and J. K. Smith, Effects of differential feedback on students' examination performance, J. Exp. Psychol. Appl. **15**, 319 (2009).

[83] L. Ketonen, Exploring Interconnections between Student Peer Assessment, JYU Dissertation, Feedback Literacy and Agency, 2021.

[84] R. A. Tejeiro, J. L. Gómez-Vallecillo, A. F. Romero, M. Pelegrina, A. Wallace, and E. Emberley, Summative self-assessment in higher education: Implications of its counting towards the final mark., Electron. J. Res. Educ. Psychol. **10**, 789 (2012).

[85] J. H. Nieminen, Disrupting the power relations of grading in higher education through summative self-assessment, Teach. High. Educ. **27**, 892 (2022).

[86] J. McArthur, M. Blackie, N. Pitterson, and K. Rosewell, Student perspectives on assessment: connections between self and society, Assess. Eval. High. Educ. **47**, 698 (2022).

[87] P. Black and D. Wiliam, Developing the theory of formative assessment, Educ. Assess. Eval. Accountability **21**, 5 (2009).

[88] L. Ketonen, P. Nieminen, and M. Hähkiöniemi, The development of secondary students' feedback literacy: Peer assessment as an intervention, J. Educ. Res. **113**, 407 (2020).

[89] M. Bearman, J. H. Nieminen, and R. Ajjawi, Designing assessment in a digital world: An organising framework, Assess. Eval. High. Educ. **48**, 291 (2022).

[90] R. Sembey, R. Hoda, and J. Grundy, Emerging technologies in higher education assessment and feedback practices: A systematic literature review. SSRN Electronic Journal (2023), http://dx.doi.org/10.2139/ssrn.4328075.

[91] J. H. Nieminen and L. Ketonen, Epistemic agency: The missing link between assessment (to be published).

[92] N. G. Holmes and C. E. Wieman, Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses, Phys. Rev. Phys. Educ. Res. **12**, 020103 (2016).

[93] K. Topping, Peers as a source of formative and summative assessment, *SAGE Handbook of Research on Classroom Assessment* (SAGE, Thousand Oaks, CA, 2013), p. 395.