Editors' Suggestion

Physics demonstrations, science show, or hands-on practical work? Exploring students' intrinsic motivation

Alexandr Nikitin, Petr Kácovský, Marie Snětinová,[*] Martin Chvál, Jitka Houfková, and Zdeňka Koupilová

Department of Physics Education, Faculty of Mathematics and Physics, Charles University, Czech Republic

(Received 3 January 2025; accepted 14 April 2025; published 6 May 2025)

In this quantitative study, we explored learners' intrinsic motivation towards three different experimental physics educational activities offered by our department to upper-grade high school students. The inspected activities were lecture demonstrations, a science show, and hands-on practical work. The respondents were Czech high school students ($N = 9647$) who participated in these activities. From each of these activities, a highly comparable subset of 1128 respondents was selected for the comparative analysis. To assess their intrinsic motivation and its covariates, we used three dimensions of the Intrinsic Motivation Inventory: *interest/enjoyment*, *effort/importance*, and *value/usefulness*. The data were evaluated using a unidimensional IRT Graded Response model. Our study revealed that the science show was more appealing to high school students than traditional lecture demonstrations and, surprisingly, even more appealing than practical lab work. This increased interest was more pronounced among students who identified themselves as less diligent in physics. Paradoxically, the respondents perceived the science show as less useful than the other two activities. Delving deeper, gender differences in perceived intrinsic motivation were generally minor, with one exception: in a comparison of the interest generated by the science show and practical work, girls favored the science show more than boys did. These findings underscore the need for a nuanced understanding of pedagogical strategies and offer insights into different approaches to promoting student engagement and interest in physics. Our research could assist educators in distinguishing the activities that might be more beneficial in terms of generating interest among particular groups of students.

DOI: 10.1103/PhysRevPhysEducRes.21.010146

I. INTRODUCTION

Various roles are (often uncritically) attributed to experiments, including promoting conceptual understanding, stimulating interest, illustrating scientific methods, developing manual skills, and fostering an understanding of the nature of science. In this study, we focus solely on the motivational aspects of physics experiments at the high school level. Substantial evidence indicates that physics experiments can generate intrinsic motivation and situational interest. However, this raises questions that have not yet been adequately explored, such as how effective different approaches to experimenting (traditional lecture demonstrations, science shows, hands-on practical work) are in promoting interest or how students' attitudes towards school physics influence their preferences for these various approaches. The primary objective of this study is to explore the answers to these questions.

II. INTRINSIC MOTIVATION AND INTEREST

Intrinsically motivated people perform activities for their inherent satisfaction and enjoyment rather than being driven by external rewards. In the second half of the 20th century, cognitive evaluation theory became one of the leading frameworks for analyzing the relationship between extrinsic and intrinsic motivation [1], a theory that assumes that human behavior is driven by two essential psychological needs: competence and autonomy. Over time, as empirical evidence concerning personality development and human behavior has accumulated, cognitive evaluation theory has been gradually supplemented by other associated mini-theories, resulting in the emergence of self-determination theory [2,3]. Currently, this metatheory comprises six mini-theories, addressing the social contexts of intrinsic motivation or the issues of well-being, life goals, or relationship quality [4]. The two original basic psychological needs (competence and autonomy) have been supplemented with a third one—relatedness or the need to feel socially connected [5].

[*]Contact author: marie.snetinova@matfyz.cuni.cz

A. Interest and its development

The closely related concept of interest is often addressed in relation to intrinsic motivation. Some researchers conceive interest as a psychological state or an individual disposition, sometimes both [6,7], or as a set of 'motives that may result in intrinsically motivated behavior' [6]. Other authors [8–10] describe interest as an emotion. Regarding intrinsic motivation, interest is considered its implicit aspect [11], a form of preconditioning [12], or an antecedent [9]. Reeve [8] proposes that interest is one of the two drivers of intrinsic motivation, the other being enjoyment. While interest contributes to focusing attention on one activity, enjoyment maintains the determination to persist. The conceptual proximity between interest/enjoyment and intrinsic motivation is essential in psychological measurements; participants' expressed interest/enjoyment is considered a measure of intrinsic motivation [13].

A strong rationale for promoting interest in schools is its impact on learning gain. Although some reviews do not show such an effect [14], prevailing conclusions suggest that interest acts as a mediator of learning [15], can increase learning gain, and predict school performance [16,17], and this is an important reason why students choose challenging tasks [18].

It is widely accepted that the process of interest development has two characteristic stages—situational and individual. While situational interest is a temporary affective reaction generated by situations that attract attention [6,9,19], individual interest is a relatively stable and persistent orientation towards certain activities or subject areas [9]. More detailed models of interest assume that each of the concepts of situational and individual interest has its own structure. Hidi and Renninger [19] suggest a four-stage model of interest development, in which situational interest has an initial triggering phase—a short-term attention increase—and a subsequent maintenance phase, where attention remains focused for a longer time. Similarly, individual interest is assumed to have two phases, emerging and well-developed, both of which are already characterized by the conscious effort to engage in certain activities. Over time, competing interest development models have been described [9], challenging the linearity of the four-stage model (i.e., that earlier phases of interest are preconditions for higher levels). Schiefele [9] suggests that individual interest can be supported in two separate ways—by triggered and maintained situational interest. Rotgans and Schmidt [14] note that previously developed individual interest contributes to situational interest when novel stimuli emerge.

B. Triggering interest in science education

Triggering situational interest is a crucial initial phase in interest development, from which the subsequent stages unfold. Activities aimed at triggering interest involve challenge, novelty, surprise, and complexity [16]. In a science classroom, such activities encompass the following: active learning methods [20], collaborative and team-based learning [9,19,21], problem-based tasks [16], hands-on practical and laboratory work [22–26], lecture demonstrations [24,25,27], or projects and fieldwork [9,16,19].

In this study, we compare three activities that represent three different approaches to integrating experiments into physics education: traditional lecture demonstrations, a science show, and hands-on practical work. In the next section, we describe these activities and formulate our research questions.

III. RESEARCH CONTEXT: THREE DIFFERENT ACTIVITIES TO BE COMPARED

Presenting physics experiments and providing access to them for upper-grade high school students have a decades-long tradition at the authors' workplace. During the last ten years, this has been mainly conducted through three activities (the subject of our research): performances known as *Physics Demonstrations For Upper Secondary Students* (DEMOs), a science show *Physics Through All Senses* (PTAS), and hands-on experimenting in the *Interactive Physics Laboratory* (IPL).

All three activities serve educational goals, such as fostering understanding, and objectives aimed at engaging a broader audience, such as triggering interest in physics; however, each emphasizes these goals differently. The common aspect among these activities is that they are led by experienced teacher trainers from the authors' workplace or well-trained pre-service physics teachers, all of whom exploit a wide range of experimental tools and share their experiences. Four of the authors participate, or have participated, in organizing these activities.

Each of the three activities has its specific characteristics, which are outlined below. For quick reference, a brief comparison of the activities is provided in Table I.

A. Physics demonstrations for upper secondary students

The project "Physics Demonstrations For Upper Secondary Students" (DEMOs) [27], which involves physics presentations with a structured sequence of carefully selected experiments, has a tradition of more than 30 years at the authors' workplace. Through a thoughtfully chosen series of intriguing demonstration experiments, each theoretically explained and focused on a specific physics topic, students are guided toward a deeper understanding of that topic, develop insights into the connections and relationships between subtopics, and learn the theoretical concepts that explain the physics phenomena observed.

Currently, there are seven distinct demonstrations, each aimed at a different physics topic. They take place in a university lecture hall and are performed by one or two lecturers who engage students through questions and occasional participation in demonstrations. The formal

TABLE I. Comparison of selected aspects of examined activities. Note: Numerical values refer to a single instance of each activity.

	Physics Demonstrations For Upper Secondary Students (DEMOs)	Physics Through All Senses (PTAS)	Interactive Physics Laboratory (IPL)
Experimental activity	lecture demonstrations	science show	hands-on small-group practical work
Location	university, out-of-school	students' school	university, out-of-school
Duration	75 min	90 min	120 min
Emphasis on understanding	considerably high, lecturers provide both simple and advanced explanations	lowest, the emphasis is on entertainment, lecturers provide relatively brief explanations	highest, students formulate their explanations that are refined by lecturers
Topics per one instance of activity	monothematic	experiments across various topics	monothematic
Character of typical equipment	a wide range of advanced equipment	mainly low-cost or everyday equipment	a wide range of advanced equipment
Lecturers	teacher trainers	mainly preservice physics teachers	teacher trainers, Ph.D. students, senior pre-service physics teachers
Typical number of participants	60–90 students	30–60 students	10–16 students
Level of student engagement	low, only a few volunteers	medium, more frequent engagement of students	high, independent students' work
Structure of typical student groups	whole classes, sometimes classes from different schools together, and occasionally even several classes from different years together	whole classes, even multiple classes from different years together, always students from the same school	students from one class of one school, sometimes a selection of students (from elective physics seminars)

setting and structured delivery emphasize the systematic exploration of physics principles.

B. Physics through all senses

"Physics Through All Senses" (PTAS) [28] is a project that has been running for more than 10 years and features approximately 25 engaging physics demonstrations inspired by the topic of the human senses. The aim of the show is to stimulate students' interest and foster students' conceptual (qualitative) understanding of selected aspects of physics. The tools and equipment chosen are deliberately simple and often inexpensive, enabling students to replicate some experiments at home.

The show is presented by two lecturers, who are pre-service physics teachers from the authors' workplace. The relatively small age gap between the lecturers and the students helps create a more informal and relatable learning atmosphere. As the lecturers come to students in their schools, PTAS is the only one of the three surveyed activities that take place outside the university, making it more accessible.

C. Interactive physics laboratory

The "Interactive Physics Laboratory" (IPL) [23], established in 2008, offers students an opportunity to grasp physics through hands-on practical work. In structured inquiry settings, students work in small groups at experimental stations guided by prepared worksheets. Their work involves linking theory to the experimental setup, making predictions, conducting quantitative and qualitative experiments, and evaluating and interpreting data. Thus, the laboratory aims to provide students with an experiential understanding of experimentation as a key method of learning about the world.

Currently, there are ten experimental sets, each focusing on a single field of physics. Each set consists of four to six experimental stations. During one visit, the student groups work on stations that belong to one set pre-selected by their teacher; finally, students present findings from one of the stations to the other groups. Two lecturers are always present to assist students who may need help, ask guiding and broadening questions, and take care of the equipment.

D. Research goals and questions

The aforementioned activities illustrate three common approaches to introducing students to physics experiments. Previous studies have suggested that all three approaches can stimulate students' situational interest [22–25,27]. Following on from the discussion in the Introduction, it is relevant to investigate the relative effectiveness of these approaches. To provide a structured framework for our investigation, we formulated the following research questions related to physics experimentation at the high school level:

- *RQ1:* How does varying student engagement in experimental activities influence students' intrinsic motivation?
- *RQ2:* To what extent is the perception of activities with varying student engagement associated with gender?
- *RQ3:* Do students' self-reported diligence in and perceived difficulty of physics as a school subject affect their perceptions of the three activities equally?

The first and main research question addressed the key motivation behind our study. Existing research on experimental activities in science education tends to focus on a single experimental approach (e.g., lecture demonstrations or practical work), examining its impact on students and its benefits and limitations. However, studies comparing different experimental activities within a single context are limited. Regarding chemistry, two studies are worth mentioning. Lin *et al.* [24] studied the impact of various learning activities, including lecture demonstrations and hands-on experiments, on students' situational interest, while Thompson and Soyibo [29] investigated how lecture demonstrations, discussions, and hands-on work influence students' attitudes toward chemistry. In physics, comparisons of experimental approaches, particularly lecture demonstrations and practical work, have primarily focused on high school students' learning gains. Thijs and Bosch [30] used the topic of forces on objects at rest in their study, whereas Winkelmann and Erb [31] focused on geometric optics. Further, Kestin *et al.* [32] have explored the effectiveness of live and online lecture demonstrations in an introductory university course. Nevertheless, there are no large-scale studies directly comparing the effects of lecture demonstrations and practical work on interest (intrinsic motivation) in high school physics.

In the complementary RQ2, we aim to contribute to an understanding of which activities can narrow (or at least not strengthen) existing gender differences. Research on the gender gap in experimental activities in physics has predominantly focused on practical work, with somewhat contradictory findings. While practical work is recognized for its potential to empower girls [33,34], it is also associated with higher perceived pressure and lower feelings of competence among girls compared to boys [23,35]. The literature has neglected the issue of potential gender disparities in the perception of lecture demonstrations or science shows, an aspect we aim to address.

The complementary RQ3 was driven by findings of Kácovský and Snětinová [27] on lecture demonstrations, which identified self-reported diligence in physics being closely related to some 'general like of physics', suggesting its robust predictive power for students' intrinsic motivation. The same authors also found that a higher perceived difficulty of school physics supports intrinsic motivation in lecture demonstrations—it (maybe paradoxically) serves as a driver of interest rather than a barrier [27]. Building on these findings, we aim to scrutinize whether these effects exhibit comparable strengths across all three activities, with varying student engagement under examination.

IV. METHODOLOGY

This study employed a quantitative research approach, addressing high school students aged 15–20 who participated in the previously described experimental activities.

A. Participants

Nearly 10,000 students participated in this study. Most of them were attending general selective (grammar) schools that prepared them for university studies. DEMOs and the IPL were visited mainly by students from the capital city and its surrounding areas, whereas PTAS was almost exclusively conducted outside the capital city.

Typically, whole classes attended DEMOs and PTAS, corresponding to a sample of participants reasonably representative of the high school population. In the case of IPL, however, the students visiting the lab are more often from elective high school physics seminars than in the other two activities. This group of students would therefore be less representative.

To ensure comparability in subsequent analyses, a subset of 1128 students was selected from each activity. This selection was made to balance key characteristics across groups, which are described in detail in Sec. IV C 1.

B. Data collecting: Tools and procedures

To assess the students' perceptions of each activity, a Questionnaire on Intrinsic Motivation (QIM; [36])[1] was designed, comprising parts labeled A, B, and C. QIM-A collected basic data on the students: gender, age, year of study, physics grade, and intention to study science, technology, engineering, and mathematics (STEM) at a higher education institution. QIM-B was inspired by the Czech research of Pavelková *et al.* [37] and focused on students' perception of physics as a school subject. It comprises six statements assessed on a five-point Likert scale. The first four statements explored the extent to which physics as a school subject is *popular*, *difficult*, *important*, and *interesting* for students, and the last two statements identified how *gifted* and *diligent* the students feel they are in physics.

QIM-C is based on the Intrinsic Motivation Inventory (IMI), a multidimensional questionnaire grounded in self-determination theory designed to assess participants'

[1]In previous research [23,27], this questionnaire was labeled differently. Specifically, the DEMOs activity was labeled DIMI, and the IPL activity was labeled LIMI. All three versions of the questionnaires can be found in [36].

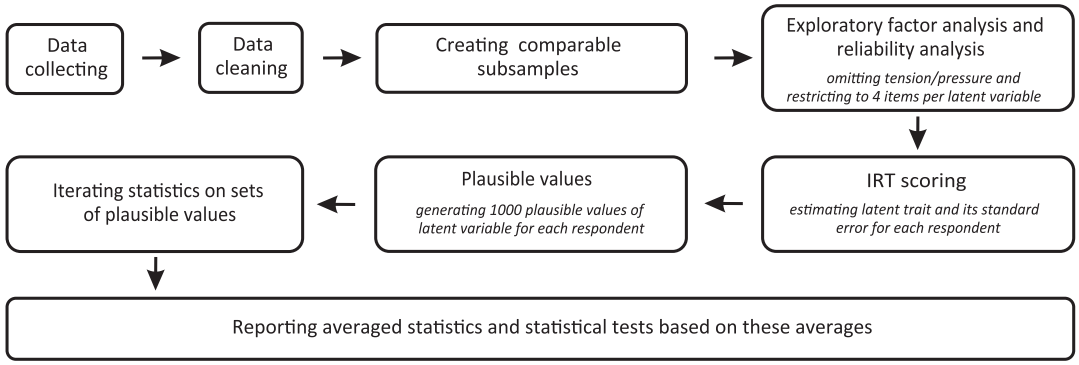

FIG. 1. Diagram of the data processing procedure.

subjective perception of an activity [13]. The IMI offers a choice of seven dimensions, with each dimension represented by several statements rated on a seven-point Likert scale. For the purpose of our research, we selected four dimensions: *interest/enjoyment* as an inherent dimension and its associated dimensions, including *effort/importance*, *felt pressure/tension*, and *value/usefulness*. The first three dimensions are represented by four items, whereas the last dimension comprises six items.[2]

The subsequent explanation of the dimensions used in this study is adapted from Monteiro *et al.* [38]. The *interest/enjoyment* dimension evaluates a person's level of interest in and pleasure derived from an activity. The *effort/importance* dimension reflects the degree to which individuals invest their capacity in an activity. The *value/usefulness* dimension encompasses the idea that individuals are more inclined to internalize and develop self-regulatory activities when they perceive an experience as valuable and useful. The *felt pressure/tension* dimension assesses the extent to which participants experience pressure to succeed in a particular task.

Data collection took place in the years 2017–2018 for DEMOs, 2018–2020 for PTAS, and 2017–2020 for IPL. The QIM was administered by the activity facilitators as a paper-and-pencil questionnaire involving data collection from 37 schools for DEMOs, 28 schools for PTAS, and 23 schools for the IPL. The students completed the questionnaire anonymously immediately after each activity. They were informed that by completing and returning the paper questionnaire, they were giving consent for data processing. Although no specific time limit was imposed on the respondents, the majority completed the questionnaire within 5 min. Researchers were present at the administration of QIM only when they were the person facilitating the activity. In cases where a researcher was not present, the facilitator received instructions on how to administer the QIM to ensure comparable conditions.

[2]The IMI version for the IPL also included a fifth dimension, perceived competence, along with one extra item in the value/usefulness dimension. However, owing to this study's focus on comparing activities, these specific data are not addressed further.

C. Data analysis

After collection, data from the paper questionnaires were digitized using FormScanner software and analyzed using IBM SPSS Statistics 26, IRTPRO 6.0, and MS Excel. The data analysis process, as depicted in Fig. 1, is aligned with the structure described in this section. In this study, a unidimensional IRT analysis with prior factor analysis was used.

1. Data filtering and subsets of comparable students

Evidently, invalid responses were filtered out from the raw data using pairs of direct- and reverse-worded QIM-C items (DRIs: three pairs for DEMOs and PTAS and four pairs for the IPL). Respondents were excluded from further analyses when at least one of the following criteria was met: (a) in one pair of DRIs, the difference in score was equal to 6; (b) in at least two DRIs, the difference in score was > 5. This led to the removal of 107 respondents (2.2%) from the DEMOs data, 104 (3.5%) from the PTAS data, and 25 (1.2%) from the IPL data. The data remaining after filtering are referred to as *full data*.

Owing to the distinct nature of DEMOs, PTAS, and IPL activities and their respective audiences, we created a subset of respondents for comparative analysis. This *subset* was composed of trios of respondents each of whom (a) participated in a different activity and (b) provided identical answers to the following QIM-A and QIM-B items: *type of school*, *gender*, *year of study*, *intention to study STEM, giftedness,* and *diligence*. Basic descriptions of *full data* and *subset* are presented in Table II.

By selecting respondents in this manner, the dataset was inherently biased by the activity with the fewest respondents, the IPL, which is often visited by students also attending elective physics seminars. Thus, the students in the *subset* were, on average, more inclined toward physics than those in the *full data* and perceived themselves as

TABLE II. Basic descriptive statistics for full data (gray cells) and the subset (white cells).

	Activity	No. of responses	Gender (%)		Year of study (%)				School (%)	
			Female	Male	1st	2nd	3rd	4th	Grammar	Vocational
Full data	DEMOs	4808	53	45	32	33	28	7	96	4
	PTAS	2840	55	43	32	37	29	1	68	32
	IPL	1999	37	60	25	38	28	9	87	13
Subset	1128 respondents for each activity		49	51	23	44	31	2	93	7

TABLE III. Differences between students selected for individual activities, full data (gray cells) and the subset (white cells). Note: All the QIM-A and QIM-B items are coded reversely compared to previous studies [23,27]. This has no effect on the QIM itself; it is conducted only to make numerical interpretations of the data clearer.

Statements	Range	DEMOs mean (SD)	PTAS mean (SD)	IPL mean (SD)	Full data mean (SD)	Subset mean (SD)
I intend to study STEM at university	1: definitely not 5: definitely yes	3.02 (1.45)	2.91 (1.47)	3.50 (1.36)	3.09 (1.46)	3.43 (1.37)
I intend to study physics at university	1: definitely not 5: definitely yes	1.63 (0.89)	1.66 (0.91)	2.04 (1.06)	1.72 (0.95)	1.92 (1.03)
Importance of physics for me	1: unimportant 5: very important	2.95 (0.98)	2.89 (0.97)	3.35 (0.99)	3.01 (0.99)	3.22 (0.96)
My giftedness in physics	1: talentless 5: very gifted	3.11 (1.00)	3.08 (1.04)	3.46 (0.93)	3.17 (1.01)	3.39 (0.90)
My diligence in physics	1: lazy 5: very diligent	3.34 (1.03)	3.31 (1.05)	3.59 (1.02)	3.39 (1.04)	3.60 (0.92)

more gifted and diligent in physics. This composition explains why the *subset* resembles the IPL group more closely than the other activities (a pattern also visible in Table III).

To better understand these differences, in Table III, we examined the responses in QIM-A and QIM-B items across the individual activities, *full data*, and the *subset*. We note that for each statement, the absolute value of Cohen's d (between *full data* and the *subset*) ranged between 0.21 and 0.24, indicating an effect size that is "not so small as to be trivial" [39]. Being aware of these differences, we performed all statistical analyses twice: once for *full data* and once for the *subset*.

2. Factor structure

The QIM-C *full data* were analyzed using exploratory factor analysis (KMO $= 0.92$; Bartlett's test of sphericity significance with $p < 0.001$) and, more precisely, using principal axis factoring [40,41], rotating the solution using a direct Oblimin rotation. The results revealed a four-factor structure, with each factor in accordance with one of the IMI dimensions. The factor associated with the *felt pressure/tension* dimension is virtually uncorrelated with the other factors (correlation coefficients ranging from 0.08 to 0.12), while the correlation coefficients between the remaining factors fall in a range from 0.50 to 0.70.

The internal consistency of the data for each IMI dimension was assessed using Cronbach's alpha. This analysis of reliability showed that the *felt pressure/tension* dimension behaves poorly—showing a low value for the alpha coefficient (0.54) and a correlation matrix far from τ-equivalent with items hardly correlating with each other (r values between 0.11 and 0.37). Omitting *felt pressure/tension* raises the overall alpha for the whole QIM-C from 0.86 to 0.91, similar to a previous study by Kácovský and Snětinová [27]. Furthermore, two items from the *value/usefulness* dimension most deviating from τ-equivalence were omitted, reducing all three remaining IMI dimensions to four items (lowering the overall alpha coefficient from 0.91 to 0.89).

As the reliability analysis of the IMI dimensions in the *subset* yielded similar results, we do not discuss it here. The alpha coefficients for the dimensions, after omitting one dimension and the two items mentioned earlier, were as follows: *interest/enjoyment* (*full data* 0.87; *subset* 0.87), *effort/importance* (0.76; 0.77), and *value/usefulness* (0.80; 0.89). The resulting Cronbach's alpha value for the whole QIM-C is 0.89 for the *full data* and the *subset*.

3. Item response theory and person scoring using plausible values

IRT was employed to obtain more accurate estimates of the latent person IMI score (the person parameter in IRT terminology) for each IMI dimension. As the four items constituting each dimension are seven-point Likert scales, the unidimensional IRT Graded Response Model (as proposed by Samejima [42]) was used for person IMI score estimation using IRTPRO 6.0. The scale of IMI scores was set to a Z score (population distribution). A simulation study by Dai *et al.* [43] suggests that, due to the low missing rate ($<2\%$ for each item, 0.7% on average) and larger sample size (> 3300 in *subset* and > 9500 in *full data*), person IMI score estimates should be considerably robust. Item parameter estimation was not considered important in this study.

The values of respondents' IMI scores arising from IRT estimation should be independent of the choice of items or the characteristics of the respondent group [44]. Correlation coefficients (Pearson's) between corresponding IMI score estimates based on *full data* and *subset* are ≥ 0.999, verifying that the latent trait estimates (at least in our case) were independent of the respondent group. Thus, only one set of IMI score estimates for each dimension was used further (the one from *full data*).

The interpretation of the resulting latent IMI scores in this study is as follows: a greater value corresponds to a higher level of respondents' identification with the corresponding IMI dimension.

Each respondent was assigned 1000 plausible values (PVs) randomly selected from a normal distribution, with a mean equal to the estimate of the respondent's latent IMI score and a standard deviation equal to the score's standard estimation error. These PVs were generated separately for each IMI dimension; thus, 1000 sets of PVs of respondents' IMI scores were created for each dimension. The individual sets of PVs generated in this manner exhibited a standard normal distribution.

4. Subsequent data analysis

Each statistic was calculated separately for each distinctive set of PVs, and the average value of this statistic is reported. The significance of the corresponding statistical test was inferred from the averaged statistics. For example, to estimate the difference between two groups, the mean, standard deviation, degrees of freedom, t statistic, and Cohen's d were calculated for each set of PVs (1000 times), and the values of these statistics were averaged; the p-value calculation was then based on these averaged values. This methodology is similar to the one used in the PISA assessment [45].[3]

First, for RQ1 and RQ2, the group descriptors were inspected by estimating the group mean, its standard deviation, and the 95% confidence intervals of the mean (assuming a standard normal distribution, thus using a critical Z value of 1.96).

To describe the differences between the activities (RQ1), we compared the IMI scores of students attending these activities using an independent-sample t test. The answer to RQ2 was handled similarly, except that the respondents were split according to *gender*, and the differences between activities were inspected separately for each gender using an independent-sample t test.

As the standard deviation in all reported groups is roughly equal to 1 (ranging from 0.9 to 1.1), the difference in group means is approximately equal to Cohen's d [46] and can therefore be considered a measure of the standardized effect size itself. Thus, Cohen's d's are not stated for t tests in the description of the results for RQ1 and RQ2 (and standard deviations are not reported in the results).

According to Cohen [46], effect sizes are small when $d = 0.2$, medium when $d = 0.5$, and large when $d = 0.8$. However, the use of these conventional thresholds has been widely discussed [47,48] because their interpretation is field specific and varies even across subdisciplines of a field. Thus, the effect sizes are commented upon only in comparison with each other [48].

A regression analysis was used to explore the relationship between students' IMI scores, their self-reported *diligence* in physics, and the perceived *difficulty* of physics (RQ3). A simple ordinary least squares linear regression was used to estimate the parameters (slope and intercept with their standard errors) of a linear model in which *diligence* and *difficulty* were treated as independent variables, and respondents' latent IMI scores were treated as dependent variables.

5. Correlations of IMI scores, diligence, and difficulty

Before discussing the results, let us address the relationships among the inspected IMI scores, self-reported student diligence, and the perceived difficulty of physics. As *value/usefulness* and *effort/importance* can be considered covariates of *interest/enjoyment*, the answers to the RQs regarding the different dimensions are somewhat interrelated. The strength of this relationship can be examined using the correlation coefficients between the IMI scores, as indicated in Table IV. These correlations affect the subsequently discussed results, which were analyzed separately for each dimension.

[3]Imputation variance is ignored due to the large number of PVs used. Sampling variance is not an issue for this study.

TABLE IV. Pearson's correlations (all significant at $p = 0.001$) between latent IMI scores and self-reported diligence and perceived difficulty. Note: Gray cells = full data, white cells = subset.

	Interest/enjoyment	Effort/importance	Value/usefulness	Diligence	Difficulty
Interest/enjoyment		0.48	0.54	0.33	0.16
Effort/importance	0.47		0.38	0.24	0.23
Value/usefulness	0.53	0.37		0.30	0.15
Diligence	0.30	0.21	0.25		0.19
Difficulty	0.14	0.20	0.12	0.16	

Notably, students' self-reported diligence is more strongly related to the IMI dimensions than to the perceived difficulty of physics as a school subject.

Table IV suggests that the results may be similar for *full data* and *subset*. Therefore, the differences between *subset* and *full data* are primarily discussed in relation to RQ1. For RQ2 and RQ3, we report only the results from the comparable *subset* because there were no substantial differences between *full data* and *subset*.

V. RESULTS

Before interpreting the results, we highlight several important considerations relevant to their interpretation. The PTAS activity consistently delivers the same science show, whereas DEMOs and IPL cover a range of topics, making comparisons across activities challenging. Additionally, in DEMOs, each topic is presented by a specific lecturer, while lecturers alternate in IPL and PTAS. However, this study does not analyze the potential effects of lecturer personality or topic selection.

Differences in student demographics also present a limitation. Students in DEMOs and IPL predominantly come from the capital, whereas PTAS participants are mostly from other regions. However, physics instruction in Czech secondary schools is relatively uniform across the country, so regional origin is unlikely to introduce substantial bias. Additionally, IPL had a higher percentage of male students, likely reflecting its appeal among those attending elective physics seminars. To mitigate the impact of these demographic differences, a subset of students with comparable characteristics was created. This allows for more valid comparisons across activities, although not for independent assessment of each group.

Finally, a higher IMI score corresponds to higher identification with the respective IMI dimension. The mean difference approximates the effect size, and significance levels correspond to independent-sample t tests, assuming unequal variances.

A. Differences between activities (RQ1)

From the simple descriptive statistics in Table V, we can infer that the students' perceptions of their engagement varied depending on the activity they participated in. This is further confirmed by the results of independent-sample t

TABLE V. Descriptive statistics of latent IMI scores split by activity. Note: Gray cells = full data, white cells = subset.

	Interest/enjoyment			Effort/importance			Value/usefulness		
	DEMOs	PTAS	IPL	DEMOs	PTAS	IPL	DEMOs	PTAS	IPL
N	4808	2840	1999	4808	2839	1999	4808	2840	1999
Mean	−0.11	0.17	0.02	−0.19	0.06	0.38	0.01	−0.13	0.17
95% C.I.	±0.03	±0.04	±0.04	±0.03	±0.04	±0.04	±0.03	±0.04	±0.04
N	1128	1128	1128	1128	1128	1128	1128	1128	1128
Mean	−0.01	0.19	0.04	−0.10	0.08	0.42	0.12	−0.09	0.17
95% C.I.	±0.06	±0.05	±0.06	±0.06	±0.05	±0.05	±0.06	±0.06	±0.06

TABLE VI. Mean differences in latent IMI scores between activities. Note: Gray cells = full data, white cells = subset. Results with $p < .01$ are denoted by *.

Interest/enjoyment			Effort/importance			Value/usefulness		
DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL	DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL	DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL
−0.28*	0.15*	−0.13*	−0.26*	−0.32*	−0.57*	0.14*	−0.30*	−0.16*
−0.21*	0.15*	−0.05	−0.18*	−0.34*	−0.52*	0.21*	−0.26*	−0.04

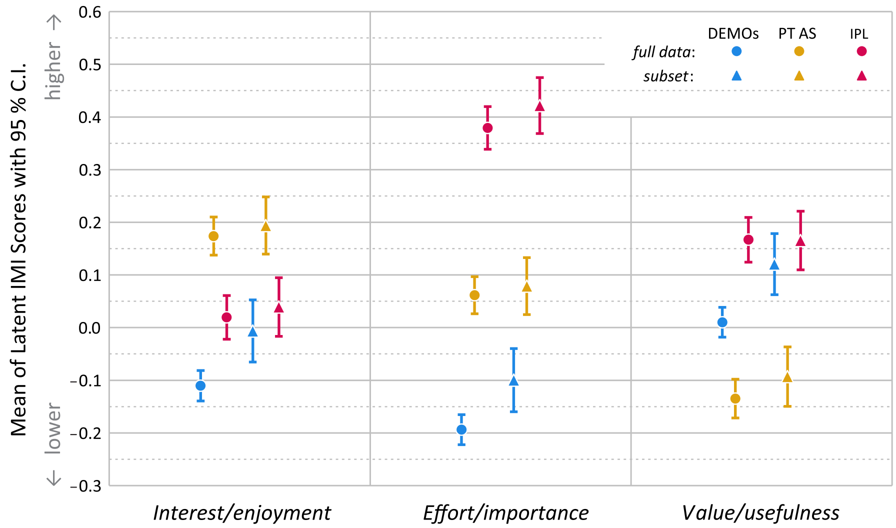

FIG. 2. Means of latent IMI scores (split by activity, with 95% confidence intervals).

tests stated in Table VI, which verify that most of the differences are significant at a 0.01 level.

While PTAS triggers the highest *interest/enjoyment* and lowest *value/usefulness* of the three activities, the situation in these two dimensions is more complicated for DEMOs and the IPL. Although DEMOs trigger lower *interest/enjoyment* and *value/usefulness* than the IPL for *full data*, the differences between these two are insignificant for the comparable *subset*.

Plotting the descriptive statistics for the *subset* and *full data* into Fig. 2 shows that hands-on experimenting in the IPL leads to significantly higher scores for *effort/importance* from students than the lecture-based approach of PTAS or DEMOs, with DEMOs requiring even less *effort/importance* than PTAS. This shows that the increase in *effort/importance* between DEMOs and PTAS is smaller than that between PTAS and the IPL.

As mentioned above, *full* data were omitted from the results, as we intended to compare the perceptions of these activities by comparable sets of students. Thus, the following data relate to the comparable *subset*.

B. Gender differences (RQ2)

As was the case with previous research findings [23,27], no significant gender differences in perceptions of DEMOs and the IPL were discovered, and this study shows that the same applies to PTAS as well. As no gender differences were observed within the individual activities, the focus moved to the differences between the three activities.

As is apparent from Table VII, the differences between the activities for female and male students were similar to those observed in RQ1 (see Table V). However, the visualization of these data in Fig. 3 reveals that the mean differences between the activities are affected by gender.

Most notably, male and female students diverged in the comparison between PTAS and the IPL regarding interest/enjoyment, with the mean difference being significant for female students (in favor of PTAS) and insignificant for male students (see Table VIII). Moreover, the effect size of this difference was substantially larger for female students, favoring PTAS over DEMOs and the IPL.

Regardless of gender, the highest *effort/importance* is attributed to the IPL and the lowest to DEMOs. The results (Table VIII) show that all the differences in *effort/importance* are statistically significant at a 0.01 level. A careful inspection of the mean differences in Table VIII suggests that the increase in *effort/importance* devoted to hands-on experimentation in the IPL is larger for female than for male students.

TABLE VII. Descriptive statistics of latent IMI scores split by gender and activity for subset.

		Interest/enjoyment			Effort/importance			Value/usefulness		
		DEMOs	PTAS	IPL	DEMOs	PTAS	IPL	DEMOs	PTAS	IPL
Female ($N = 550$)	Mean	0.03	0.26	0.00	−0.12	0.07	0.45	0.16	−0.03	0.20
	95% C.I.	±0.08	±0.08	±0.08	±0.09	±0.08	±0.08	±0.08	±0.08	±0.08
Male ($N = 578$)	Mean	−0.04	0.13	0.08	−0.08	0.09	0.39	0.08	−0.16	0.14
	95% C.I.	±0.08	±0.08	±0.07	±0.08	±0.08	±0.07	±0.08	±0.08	±0.08

FIG. 3. Means of latent IMI scores for comparable subset (split by activity and gender, with 95% confidence intervals).

Concerning *value/usefulness*, there were no significant gender differences between the IPL and DEMOs in this *subset*. Both genders attributed the lowest *value/usefulness* to PTAS.

C. Effect of students' diligence and perceived difficulty (RQ3)

The interpretation of the coefficient b in the linear regression model

$$(\text{IMI score}) = a \cdot (\text{diligence}) + b$$

is similar in nature to the differences between activities that have already been inspected in RQ1. Thus, it is not given further attention here, and the primary focus is on the slope coefficient a. The same applies to the perceived *difficulty* of physics as a school subject.

The positive values of the slope coefficients in Table IX show a general trend of increasing IMI scores with increasing

TABLE VIII. Independent-samples t tests for mean differences in latent IMI scores between activities split by gender. Note: Results with $p < 0.01$ are denoted by *.

	Interest/enjoyment			Effort/importance			Value/usefulness		
	DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL	DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL	DEMOs vs PTAS	PTAS vs IPL	DEMOs vs IPL
Female	−0.23*	0.26*	0.03	−0.19*	−0.38*	−0.57*	0.19*	−0.22*	0.03
Male	−0.17*	0.05	−0.13	−0.17*	−0.30*	−0.47*	0.24*	−0.29*	0.05

TABLE IX. *Estimated parameters of the linear regression of* diligence *and* difficulty *on IMI scores for the* subset.

		Interest/enjoyment			Effort/importance			Value/usefulness		
		DEMOs	PTAS	IPL	DEMOs	PTAS	IPL	DEMOs	PTAS	IPL
Diligence	N	1128	1128	1128	1128	1128	1128	1128	1128	1128
	a	0.34	0.26	0.36	0.25	0.20	0.22	0.27	0.24	0.28
	(S.E.)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	R^2	0.10	0.06	0.12	0.05	0.04	0.05	0.06	0.05	0.07
Difficulty	N	1122	1126	1127	1122	1126	1127	1122	1126	1127
	a	0.17	0.12	0.15	0.22	0.15	0.24	0.13	0.14	0.09
	(S.E.)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	R^2	0.03	0.02	0.02	0.04	0.02	0.06	0.02	0.02	0.01

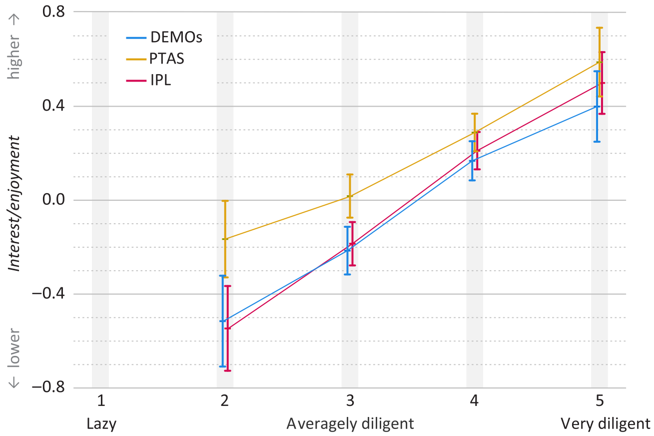

FIG. 4. Means of latent interest/enjoyment for subset with 95% C.I.s split by the respondents' self-reported diligence.

student *diligence* and increasing perceived *difficulty* of physics. In other words, students who perceive themselves as more *diligent* or who find physics more *difficult* tend to perceive greater *interest/enjoyment*, feel higher *value/usefulness*, and invest more *effort/importance* in any of the three experimental activities.

The effect of *diligence* on *interest/enjoyment* triggered by PTAS is the weakest among the three activities. *Diligence* affects *value/usefulness* and *effort/importance* similarly during all three activities.

The effect of perceived *difficulty* on *interest/enjoyment* and *value/usefulness* seems to be unaffected by the experimental approach, whereas the effect on *effort/importance* is lower in PTAS than in DEMOs and the IPL.

Further inspection of the slope coefficients in Table IX reveals that, for all three activities, self-reported *diligence* is a stronger predictor of *interest/enjoyment* and *value/usefulness* than is the case for the perceived *difficulty* of physics.

As anticipated from the nature of these data, the determination coefficients R^2 reported in Table IX are fairly low. Nevertheless, Fig. 4, as an example of the regression of *interest/enjoyment* on self-reported *diligence*, suggests that the linear approximation of the regression is justified. The figure shows that the factual differences in *interest/enjoyment* between PTAS and DEMOs/IPL decrease with increasing *diligence* of students, pointing to a lower slope coefficient for PTAS and, thus, a less potent relationship between PTAS and *diligence*. The offset of PTAS from DEMOs/IPL was determined by the difference in coefficient b.

VI. DISCUSSION

A. Students' intrinsic motivation in activities with various levels of engagement

Students' intrinsic motivation during the three activities under examination was heavily influenced by the activity's capacity to engage them, as shown in Table I. Student engagement ranges from fairly low during the experiments performed in front of a class in DEMOs to a medium level of engagement during the PTAS science show to the highly engaging hands-on experiments in the IPL.

Regarding the first research question (RQ1), while PTAS aroused the highest *interest/enjoyment*, it was perceived as the least valuable/useful of the three activities. The high situational interest triggered by PTAS may be attributed to a "halo effect" of the appealing, funny, and flashy show, essentially pointing to its place in science education and the popularization of STEM as a way to arouse peoples' interest [49].

Additional inspection of correlations between IMI dimensions shows a slightly lower correlation between *value/usefulness* and *interest/enjoyment* for PTAS than for the other activities. This suggests that the intrinsic motivation during PTAS arises as a result of factors different from those measured by the QIM.

A comparison between DEMOs and the IPL for a comparable *subset* of students produces interesting results. As stated previously, the students constituting the *subset* are more inclined toward physics and STEM than the general population. For these selected students, no significant differences in *interest/enjoyment* and *value/usefulness* triggered by DEMOs and the IPL were discovered, indicating that those students more inclined toward physics can find a lecture demonstration as interesting and valuable as practical work.

Furthermore, the *effort/importance* dimension corresponds to what we have intuitively called "student engagement," including the fact that the differences between the two lecture-based activities (DEMOs and PTAS) are smaller than the differences between these two and hands-on practical work (IPL). The observed dominance of the IPL regarding effort is to be expected, given the higher requirements for engagement from individual students during hands-on experimentation, which is especially notable for the *subset* of comparable students.

Students perceived greater *effort/importance* in relation to PTAS than for the DEMOs. There might be several reasons for this. For example, PTAS takes place at students' home schools and is attended by smaller groups of students, which may result in lower individual anonymity and higher oversight from their teachers, forcing students to put more effort into this activity. This observation is supported by prior studies that have found that the sheer size and anonymity of large classrooms are significant barriers to student participation in classroom activities (e.g. MacGregor *et al.*, 2000, as cited [50,51]). Additionally, more PTAS experiments incorporate volunteers, even the entire audience at times, and this requirement for direct participation can lead to a perception of a higher required effort. Another source of this difference can be attributed to the high *interest/enjoyment* triggered by PTAS, in accordance with the substantial correlations between these two dimensions, essentially endorsing that the activity perceived as interesting and enjoyable is also perceived as more important and that students are willing to put in more effort.

B. Gender differences

The second research question (RQ2) focused on the extent to which gender influenced the perception of the experimental activities regarding students' intrinsic motivation and its covariates. First, we examined gender differences in the activities involved. Our findings indicate no such differences in IPL data, which aligns with Murphy's [35] conviction that "practical work can play a crucial role in combating gender differences" (p. 141). This is also consistent with the suggestions of researchers who propose narrowing the gender gap through learner-centered and problem-solving activities [34,35], both of which are part of the practical work in the IPL. We also found no gender differences in the DEMOs data. The differences in the *full data* for PTAS are statistically significant in favor of female students but with negligible effect size. This indicates that the intrinsic motivation toward experimental activities in physics is predominantly independent of gender. Therefore, we believe that such activities could be effective in mitigating the gender gap in attitudes toward physics, a discipline perceived as masculine and favoring men [52,53].

Regarding the comparison across the studied activities, the results for both genders separately follow the results from the full data. Moreover, we observed that as students' engagement increased, female students showed greater growth in expressed *effort/importance* than their male counterparts. The need for more active involvement in activities seems to elicit a stronger sense of effort in female students, possibly because they are considered more self-disciplined than male students and are more likely to suppress prepotent responses in favor of a higher goal, which requires conscious effort [54].

C. Effects of diligence and difficulty

We found that students with higher self-reported diligence in physics perceived all the activities more positively in all IMI dimensions, and the same applies to students who perceived physics as a more difficult school subject.

The results show that self-reported diligence affects the *effort/importance* and *value/usefulness* triggered by the three activities equally, suggesting that there might be a general trend connecting self-reported diligence in physics with the willingness to dedicate effort to a physics-related activity and find value in it.

For PTAS, diligence is not as strong a predictor of *interest/enjoyment* as it is for DEMOs and the IPL. PTAS triggered significantly more situational interest for less diligent students than DEMOs and the IPL, while for more diligent students, this difference was negligible. This implies that such flashy science shows are a good way to spark interest in physics among students who describe themselves as lazier in physics.

The fact that students who find physics to be a more difficult school subject tend to perceive all the activities more positively is interesting because the difficulty of physics is often mentioned as a reason for its unpopularity. For example, Williams *et al.* [55] state that there is evidence that the perception of a subject as difficult tends to result in the development of a general negative attitude toward that subject. Lavonen *et al.* [56] suggest that students who perceive physics as difficult do not try to learn it and lose interest in it quickly. The strength of the difficulty of physics as a predictor of *interest/enjoyment* and *value/usefulness* was comparable for all three presented activities; nonetheless, it was a less potent predictor of intrinsic motivation than self-reported diligence.

The difficulty of physics is a less influential predictor of *effort/importance* for PTAS than for the other two activities, suggesting that students with less positive attitudes toward physics are more likely to invest their effort in a vivid science show.

VII. LIMITATIONS

Several differences between the experimental activities themselves may have affected the results. First, the PTAS activity consistently delivers the same science show, whereas individual runs of DEMOs and IPL activities offer several different topics. Consequently, students' responses to the QIM-C were tied to a specific topic in the case of DEMOs and the IPL. Second, in the case of DEMOs, each topic is directly associated with a particular lecturer, whereas for the IPL and PTAS, the lecturers alternate. In this study, we did not delve into whether the QIM-C results depend on the specific topic or the personality of the lecturer, but both these aspects are the focus of an ongoing research project.

Another limitation arises from the fact that the teachers who enrolled their students in any of the activities tended to be more active members of the teaching community. Furthermore, the number and origin of students participating in each activity varied. Students attending DEMOs and IPL activities come predominantly from the metropolitan area of Prague, whereas PTAS activities are primarily conducted in schools outside the capital city. The possibly unrepresentative nature of the IPL groups has already been alluded to.

Finally, the IPL had a higher percentage of male students than the other two activities. This discrepancy is likely because the IPL is often attended by students from elective physics seminars who have a greater interest in physics or require physics for their university studies.

The abovementioned differences in student characteristics arising from divergent demographic profiles and encompassing gender imbalances were addressed by establishing a subset of students whose characteristics displayed high comparability. This approach allowed for a more objective comparison of all three aforementioned activities and mitigated the potential confounding effects resulting from these differences. As these subsets of students are not representative samples of DEMOs and PTAS audiences, they do not represent appropriate samples for individual assessment; only their comparison with each other is valid.

The knowledge gained from these activities by the students is not assessed in any manner. Further, notably, the respondents completed the QIM immediately after participating in the activity, and their self-reported characteristics and attitudes toward physics were influenced by their feelings at that moment in time. Finally, data could only be collected up to the onset of the COVID-19 pandemic, after which no data collection was possible.

VIII. CONCLUSIONS AND RECOMMENDATIONS

Physics is generally considered one of the least favorite school subjects and is perceived by students as challenging. Physics experiments are a suitable means of arousing interest and providing an understanding of the world. Three types of experimental activities offered to high school students by the Department of Physics Education, CUNI MFF, were presented in this study. These activities vary in the level of active student engagement and the emphasis on understanding physics.

Our research has shown that the science show *Physics Through All Senses* (PTAS) arouses higher student interest among high school students than *Physics Demonstrations For Upper Secondary Students* (DEMOs) or the *Interactive Physics Laboratory* (IPL), especially for those who self-reported themselves as less *diligent* in physics. However, the students themselves perceived watching the science show as less useful and valuable than watching DEMOs or practical work in the IPL. Nevertheless, PTAS's capacity to arouse student interest can serve as a good starting point for developing a long-lasting interest in physics. Additionally, as lecturers visit students' home schools with the show, it can reach any student, particularly helping those with greater anxiety about physics to overcome it.

DEMOs and the IPL are comparable in arousing student interest, and they are perceived as having the same value; nonetheless, the aims and instructional approaches of these activities differ significantly. The IPL is designed for students to independently explore the laws of physics in small groups under lecturers' guidance, while DEMOs serve as a way to address numerous students, providing them with insight into the demonstrated phenomena without the need for active participation.

The differences between female and male students in their perceptions of individual approaches to experimentation were negligible. However, the difference between the situational interest triggered by PTAS and the IPL was significantly larger for girls, favoring the vivid science show.

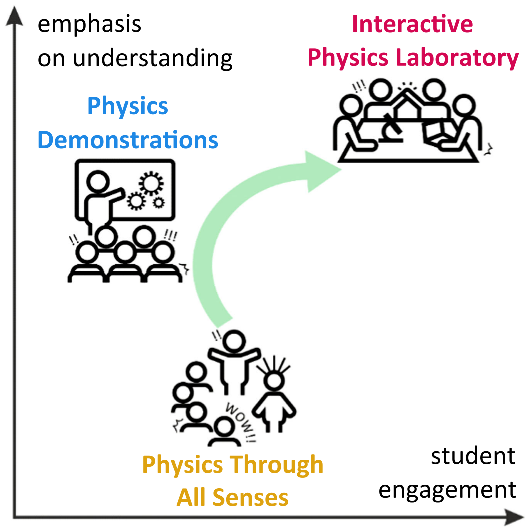

FIG. 5. Symbolic expression of the addressed experimental activities in a 2D space defined by the students' engagement and the emphasis on understanding.

From the foregoing discussion, it is evident that all three activities have a place in science and physics education. The arrow in Fig. 5 suggests the order in which the activities can be offered to students to gradually lead them from developing an interest to developing an understanding. Educators must choose an appropriate path for bringing physics to a wider range of students to reach those who are anxious about physics as well as those who seek a deeper understanding of physical phenomena.

ACKNOWLEDGMENTS

We would like to thank our colleagues from the Department of Physics Education (CUNI MFF) who helped us with data collection for this research. This work was supported by the Charles University, project GA UK No. 62223, and the Charles University Research Program COOPERATIO SOC/SSED.

A. N.: Formal analysis of study data; design of methodology; visualization and presentation of data; writing of the initial draft; critical review, commentary, and revision of the paper. P. K.: Formulation of overarching research goals and aims; management activities to produce metadata, scrub data, and maintain research; conducting a research and investigation process, data collection; oversight and leadership responsibility for the research activity planning and execution; writing of the initial draft. M. S.: Conducting a research and investigation process, data collection; management and coordination responsibility for the research activity planning and execution; visualization and presentation of data; writing of the initial draft. M. C.: Formulation of overarching research goals and aims;

design of methodology; visualization and presentation of data. J. H.: Conducting a research and investigation process, data collection; writing of the initial draft. Z. K.: Management and coordination responsibility for the research activity planning and execution; critical review, commentary, and revision of the paper.

The authors declare that they have no competing interests.

Ethical approval for this study was granted by the Research Ethics Committee, Charles University, Faculty of Mathematics and Physics on April 8, 2022 (reference number: 2022/01).

DATA AVAILABILITY

The datasets used and analyzed during the current study are openly available in the Mendeley Data repository [57].

[1] E. L. Deci, *Intrinsic Motivation* (Springer, Boston, MA, 1975), 10.1007/978-1-4613-4446-9.

[2] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior* (Springer, New York, NY, 1985), 10.1007/978-1-4899-2271-7.

[3] R. M. Ryan and E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, Am. Psychol. **55**, 68 (2000).

[4] R. M. Ryan and E. L. Deci, Brick by brick: The origins, development, and future of self-determination theory, Adv. Motiv. Sci. **6**, 111 (2019).

[5] R. M. Ryan and E. L. Deci, *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness* (The Guilford Press, New York, 2017), 10.1521/978.14625/28806.

[6] S. Hidi, An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation, in *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, edited by C. Sansone and J. M. Harackiewicz (Academic Press, New York, 2000), p. 309, 10.1016/B978-012619070-0/50033-7.

[7] A. Krapp, S. Hidi, and K. A. Renninger, Interest, learning, and development, in *The Role of Interest in Learning and Development*, edited by K. A. Renninger, S. Hidi, and A. Krapp (Psychology Press, New York, 1992), p. 3, 10.4324/9781315807430.

[8] J. Reeve, The interest-enjoyment distinction in intrinsic motivation, Motivat. Emot. **13**, 83 (1989).

[9] U. Schiefele, Situational and individual interest, in *Handbook of Motivation at School*, edited by K. Wentzel and D. B. Miele (Routledge, New York, 2009), p. 197, 10.4324/9780203879498.

[10] P. J. Silvia, *Exploring the Psychology of Interest* (Oxford University Press, New York, 2006), 10.1093/acprof:oso/9780195158557.001.0001.

[11] E. L. Deci, The relation of interest to the motivation of behavior: A self-determination theory perspective, in *The Role of Interest in Learning and Development*, edited by K. A. Renninger, S. Hidi, and A. Krapp (Psychology Press, New York, 1992), p. 43, 10.4324/9781315807430.

[12] S. Hidi and J. M. Harackiewicz, Motivating the academically unmotivated: A critical issue for the 21st century, Rev. Educ. Res. **70**, 151 (2000).

[13] Center of Self-Determination Theory, Intrinsic motivation inventory (IMI), https://selfdeterminationtheory.org/intrinsic-motivation-inventory/.

[14] J. I. Rotgans and H. G. Schmidt, Interest development: Arousing situational interest affects the growth trajectory of individual interest, Contemp. Educ. Psychol. **49**, 175 (2017).

[15] K. A. Renninger and S. Hidi, Revisiting the conceptualization, measurement, and generation of interest, Educ. Psychol. **46**, 168 (2011).

[16] J. M. Harackiewicz, J. L. Smith, and S. J. Priniski, Interest matters: The importance of promoting interest in education, Policy Insights Behav. Brain Sci. **3**, 220 (2016).

[17] O. Köller, J. Baumert, and K. Schnabel, Does interest matter? The relationship between academic interest and achievement in mathematics, J. Res. Math. Educ. **32**, 448 (2001).

[18] N. Inoue, Why face a challenge?: The reason behind intrinsically motivated students' spontaneous choice of challenging tasks, Learn. Individ. Diff. **17**, 251 (2007).

[19] S. Hidi and K. A. Renninger, The four-phase model of interest development, Educ. Psychol. **41**, 111 (2006).

[20] P. C. Blumenfeld, T. M. Kempler, and J. S. Krajcik, Motivation and cognitive engagement in learning environments, in *The Cambridge Handbook of the Learning Sciences*, edited by R. K. Sawyer (Cambridge University Press, Cambridge, England, 2006), p. 475, 10.1017/CBO9780511816833.029.

[21] L. M. Jeno, A. Raaheim, S. M. Kristensen, K. D. Kristensen, T. N. Hole, M. J. Haugland, and S. Mæland, The relative effect of team-based learning on motivation and learning: A self-determination theory perspective, CBE Life Sci. Educ. **16**, ar59 (2017).

[22] I. Abrahams, Does practical work really motivate? A study of the affective value of practical work in secondary school science, Int. J. Sci. Educ. **31**, 2335 (2009).

[23] P. Kácovský, M. Snětinová, M. Chvál, J. Houfková, and Z. Koupilová, Predictors of students' intrinsic motivation

during practical work in physics, Int. J. Sci. Educ. **45**, 806 (2023).

[24] H.-s. Lin, Z.-R. Hong, and Y.-C. Chen, Exploring the development of college students' situational interest in learning science, Int. J. Sci. Educ. **35**, 2152 (2013).

[25] D. H. Palmer, Student interest generated during an inquiry skills lesson, J. Res. Sci. Teach. **46**, 147 (2009).

[26] K. Röllke and N. Großmann, Predictors of students' intrinsic motivation in a biotechnological out-of-school student lab, Front. Educ. **7**, 859802 (2022).

[27] P. Kácovský and M. Snětinová, Physics demonstrations: Who are the students appreciating them?, Int. J. Sci. Educ. **43**, 529 (2021).

[28] P. Kácovský and V. Koudelková, Physics through all senses: Popularization of physics by using an interactive show, in *Hands-on—The Heart of Science Education*, edited by M. F. Costa, B. V. Dorrio, J. Trna, and E. Trnova (Hands-on Science Network, 2016), p. 23, http://www.ijhsci.info/hsci2016/index.php/.

[29] J. Thompson and K. Soyibo, Effects of lecture, teacher demonstrations, discussion and practical work on 10th graders' attitudes to chemistry and understanding of electrolysis, Res. Sci. Technol. Educ. **20**, 25 (2002).

[30] G. D. Thijs and G. M. Bosch, Cognitive effects of science experiments focusing on students' preconceptions of force: A comparison of demonstrations and small-group practicals, Int. J. Sci. Educ. **17**, 311 (1995).

[31] J. Winkelmann and R. Erb, Small group practical work vs. teacher demonstration in geometrical optics, in *E-book Proceedings of the ESERA 2013 Conference: Science Education Research for Evidence-Based Teaching and Coherence in Learning, Part 3*, edited by C. P. Constantinou, N. Papadouris, and A. Hadjigeorgiou (European Science Education Research Association, Nicosia, Cyprus, 2014), p. 73, https://www.esera.org/esera-2013/.

[32] G. Kestin, K. Miller, L. S. McCarty, K. Callaghan, and L. Deslauriers, Comparing the effectiveness of online versus live lecture demonstrations, Phys. Rev. Phys. Educ. Res. **16**, 013101 (2020).

[33] M. P. Freeman, The influence of laboratory instruction on science achievement and attitude toward science across gender differences, J. Women Minorities Sci. Eng. **8**, 191 (2002).

[34] P. Murphy and E. Whitelegg, *Girls in the Physics Classroom: A Review of the Research on the Participation of Girls in Physics* (Institute of Physics, London, UK, 2006), https://www.iop.org/sites/default/files/2019-04/girls-in-the-physics-classroom.pdf.

[35] P. Murphy, Gender differences in pupils' reactions to practical work, in *Teaching Science*, edited by R. Levinson (Routledge, London, UK, 1993), p. 138, 10.4324/9780203990377.

[36] Department of Physics Education, Research on intrinsic motivation in experimental activities, Retrieved from https://www.mff.cuni.cz/en/kdf/research-and-cooperation/research-on-intrinsic-motivation-in-experimental-activities (2024).

[37] I. Pavelková, A. Škaloudová, and V. Hrabal, Analýza vyučovacích předmětů na základě výpovědí žáků [Analysis of school subjects on the basis of the testimony of pupils], Pedagogika **60**, 38 (2010), https://pages.pedf.cuni.cz/pedagogika/?p=951.

[38] V. Monteiro, L. Mata, and F. Peixoto, Intrinsic motivation inventory: Psychometric properties in the context of first language and mathematics learning, Psicol. Reflexão Crítica **28**, 434 (2015).

[39] G. M. Sullivan and R. Feinn, Using effect size—or why the P value is not enough, J. Grad. Med. Educ. **4**, 279 (2012).

[40] A. B. Costello and J. Osborne, Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis, Pract. Assess. Res. Eval. **10**, 7 (2005).

[41] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, Evaluating the use of exploratory factor analysis in psychological research, Psychol. Methods **4**, 272 (1999).

[42] F. Samejima, Graded response model, in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton (Springer, New York, 1997), p. 85, 10.1007/978-1-4757-2691-6_5.

[43] S. Dai, T. T. Vo, O. J. Kehinde, H. He, Y. Xue, C. Demir, and X. Wang, Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data, Front. Educ. **6**, 721963 (2021).

[44] F. M. Lord, *Applications of Item Response Theory to Practical Testing Problems* (Routledge, New York, NY, 2008).

[45] R. Watanabe, *PISA Data Analysis Manual SPSS* (OECD, Paris, France, 2009), 10.1787/9789264056275-en.

[46] J. Cohen, *Statistical Power Analysis for the Behavioral Science*, 2nd ed. (Academic Press, New York, 1988).

[47] T. Baguley, Standardized or simple effect size: What should be reported?, Br. J. Psychol. **100**, 603 (2009).

[48] T. Schäfer and M. A. Schwarz, The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases, Front. Psychol. **10**, 813 (2019).

[49] R. Watermeyer, The presentation of science in everyday life: The science show, Cult. Stud. Sci. Educ. **8**, 737 (2013).

[50] F. Maringe and N. Sing, Teaching large classes in an increasingly internationalising higher education environment: Pedagogical, quality, and equity issues, High. Educ. **67**, 761 (2014).

[51] S. Rohi and S. Muslim, Factors influencing classroom participation: A case study of undergraduate students at education faculty, Paktia University, J. Res. Appl. Sci. Biotechnol. **2**, 99 (2023).

[52] A. J. Gonsalves, A. Danielsson, and H. Pettersson, Masculinities and experimental practices in physics: The view from three case studies, Phys. Rev. Phys. Educ. Res. **12**, 020120 (2016).

[53] *Physics Education and Gender: Identity as an Analytic Lens for Research*, edited by A. J. Gonsalves and A. T. Danielsson (Springer, New York, 2020).

[54] A. L. Duckworth and M. E. P. Seligman, Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores, J. Educ. Psychol. **98**, 198 (2006).

[55] C. Williams, M. Stanisstreet, K. Spall, E. Boyes, and D. Dickson, Why aren't secondary students interested in physics?, Phys. Educ. **38**, 324 (2003).

[56] J. Lavonen, R. Byman, K. Juuti, V. Meisalo, and A. Uitto, Pupil interest in physics: A survey in Finland, Nord. Stud. Sci. Educ. **1**) 72 (2005).

[57] A. Nikitin, Dataset for "Physics demonstrations, science show, or hands-on practical work? Exploring students' intrinsic motivation", Mendeley Data, v2 (2024), 10.17632/vfsjk42t27.2.