Learning enhancing emotions predict student retention: Multilevel emotions of Finnish university physics students in and outside learning situations

Laura Timonen[*]
Nano and molecular systems research unit, Faculty of Science, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland
and Learning and learning processes research unit, Faculty of Education, University of Oulu, P.O. Box 2000, 90014 Oulu, Finland

Kalle Juuti
Department of Education, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland

Sari Harmoinen
Learning and learning processes research unit, Faculty of Education, University of Oulu, P.O. Box 2000, 90014 Oulu, Finland

(Received 17 November 2021; accepted 17 August 2022; published 14 October 2022)

Research on student retention in higher education (HE) physics could benefit by studying emotions in the context of engagement and learning. However, popular retention theories include only a narrow selection of emotions, creating a need to look elsewhere. In this study, we borrow the lens of an affective engagement model, the framework of an optimal learning moment, which has rarely been used in HE research so far. It defines situational engagement and three categories of learning enhancing, detracting, or accelerating emotions via twelve singular situational emotions. These twelve emotions in learning and other situations form intensive longitudinal data collected from twenty students using the experience sampling method (ESM) during their first two months of studying physics in a Finnish university. A two-level hierarchical dataset consisting of ESM measures ($N_1 = 440$) and student records ($N_2 = 20$), with gender as a background factor, are analyzed in two steps: first with hierarchical linear modeling, followed by multinomial logistic regression, giving results on both levels of the hierarchy, which is quite uncommon still. The results show how situational engagement and learning situations are separately connected to situational emotions and, further, how especially the learning enhancing emotions connect to success in courses (passing, grades) and first year student retention, surpassing the effect of course success.

DOI: 10.1103/PhysRevPhysEducRes.18.020126

I. INTRODUCTION

Student retention in higher education (HE) remains a serious issue. It is estimated by the Organisation for Economic Co-operation and Development (OECD) that in its member countries (including USA and Finland) roughly a quarter of students entering any bachelor's degree program have not graduated nor are they enrolled in tertiary education after the theoretical duration of the degree plus three years [1]. HE physics, an already low degree producing field [2,3], is ideal for student retention research, as it is unbearable to continue losing roughly one-quarter of each cohort.

The role of emotions and engagement in the context of learning and especially positive emotions in academic settings would benefit from more research [4–6]. There is evidence that interest towards the subject to be learned and overall academic self-efficacy [7] and level of confidence [8] could affect retention or reasons to leave HE. Although popular student retention theories consider learning and academic achievements important [9,10], they seem to include only few emotions as part of their frameworks and often in a minor role [e.g., [11,12]]. Some engagement theories, however, consider emotions an integral part, especially the ones restricted to the psychological view of engagement, which divides engagement into behavioral, cognitive, and affective sectors [4]. It is this affective engagement perspective which we adapt in the study presented in this article, asking how situational engagement and different emotions in learning and other situations are connected to student retention.

[*]To whom all correspondence should be addressed.
laura.timonen@oulu.fi

From an institutional perspective, probably the easiest way of monitoring learning and predicting student retention is by following which students pass courses and what grades they get. Research shows that grade retention, meaning a repetition of courses, terms or even a full year, increases dropout probability [13]. Emotions such as hope and anxiety have been connected to pre-university physics course's students' motivational predictors and study outcomes (grades); hope positively and anxiety negatively [14]. Positive emotions about self, academic achievements, and study time have been shown to be more important than negative emotions, concerning university students self-regulated learning and motivation, and that the overall effect of positive emotions directly and through self-regulated learning and motivation on academic achievements (passed tests and GPA) was significant and positive [6]. Specific positive emotions, enjoyment and pride, have been connected to higher final grades, again working through self-regulated learning [15]. Recently, sense of belonging, consisting in part of emotions like feeling comfortable or worried, has also been shown to predict physics course grades [16]. We will include course success (passing and grades) as part of our study as a possible mediating variable between situational emotions and student retention.

Women might be more likely to graduate from tertiary education than men [3], but there is and has been for decades an unbalance of genders within certain science, technology, engineering, and mathematics disciplines, e.g., in physics [17]. Bean suggests that gender, ethnicity, or age would not matter in how psychological processes leading to student retention works [10]. However, there is historical evidence that women and men might leave HE for different reasons [11,18]. Previous studies indicate that women might be more susceptible to science anxiety in general science courses [19] and in physics courses [20]. A study conducted in Finnish context showed women to have more anxiety and less self-efficacy emotions than men concerning HE physics, and that low anxiety and high self-efficacy lead to staying [21]. Women have been found to report more stress than men, and vice versa for feeling happiness, confidence, and active [22]. However, women might show emotions more intensively than men [23]. Research also shows there might be gender differences in introductory physics courses which favor the males, but which are not connected to later achievements in physics [24]. All this makes gender a meaningful individual level factor and it will be included in our study as a predictor variable not only for situational emotions, but for course success and retention too.

Both gender and course success (as grades and passing) are static variables, easy to gather from student records. But emotions and student engagement are dynamic and time bound, thus it would not be enough to measure them once, but to follow them longitudinally [4]. Acknowledging situational and individual levels of engagement is suggested as important, and the results on the greatest source of variation are not unanimous [25,26]. Individual level, or between-individuals, variation refers to an individual's inclination to feel emotions in a unique scale compared to another individual. Situational level, or within-individual, variation is caused purely by the differences between measurement situations, and should ideally be separated from the differences between individuals. For this, a hierarchical dataset and suitable analysis methods are needed. It has been suggested that using hierarchical statistical methods, if possible, would be important in physics educational research [27]. For example, nonscience anxiety has been shown to be an indicator of science anxiety [19,20], and sources of stress can be tied to learning situations or wider external stressors [28], suggesting that the differences on an individual level besides situational are important. In our study, we have applied the experience sampling method (ESM) [29], in which a longitudinal data is collected from a sample of students, creating a hierarchical dataset (several responses from different situations per participant), to be later analyzed with hierachical methods, which separate the individual and situational predictors, outcomes, and variances. We will present in detail how hierarchical analysis was used to produce results not just on the lowest level of the hierarchy, which in this case are the situational level emotions, but on a higher level too, in this case the individual level course success and retention, which is still quite rare.

The article continues with a short review of emotions in popular retention theories and then introduces the framework of optimal learning moment (OLM) [22], which is the engagement model chosen as the core framework for our study. We then inform the reader briefly about the Finnish HE system and the student retention situation in Finland, as the participants of our study are a sample of Finnish university first year physics students. The perspective of our study is an institutional one and it concentrates on one education program within one university. This might seem limiting but is actually suggested when exploring student experience to bring out the individual character of that particular setting [4,18].

After the aim is summarized in three research questions, we present the study, where situational engagement and emotions, defined in the OLM, of a sample of twenty Finnish HE physics students in and outside learning situations in the first two months of their studies were collected with ESM and analyzed with hierarchical methods. We explore how situational factors such as engagement and being in a learning situation affect other situational emotions, while controlling for gender. Then we examine how these emotions are connected on the individual level to the participants first semester course passing and course grades, and student retention after the first year, again controlling for gender.

II. THEORETICAL BACKGROUND

A. Theory perspective on student retention is lacking emotions

Among student retention research, the theories and models of Astin [30], Bean [10,31], Cabrera [32,33], Pascarella [12,34], and Tinto [9,18] have been very influential [35–38]. Their theories consider processes, such as interactions, integration, and retention or drop-out, which are without a doubt emotional. However, within these theories emotions are rarely considered in detail or situated consistently in the models. Exceptions to this trend are mentions of self-efficacy, stress, satisfaction, boredom, and confidence as attitudinal, endogenous or (psychological) outcome variables [10,12,30–32]. Bean [31] suggests that the intent to leave, the last step before drop-out, would be best explained by what he calls *attitudinal variables*, including the said satisfaction, boredom, and confidence. These variables describe "the psychological results of interacting with an organization" [31] (p. 19). Bean and Eaton's later psychological model of retention [10] describes a network of psychological processes, mentioning change in stress and confidence as a result of students' coping behavior (approach or avoid). Applying this model, it has been shown that stress has indirect effects on retention through institutional commitment, intent to return, and intended academic outcomes [39]. By looking at student retention through an affective engagement framework, it is possible to explore a wider range of emotions, while sewing a thread between the student retention and engagement research lines, which have been accused of being siloed [35].

B. Framework of optimal learning moment connects situational emotions within learning situations

An example of an affective engagement model, chosen as the base for our study, is the framework of optimal learning moment [22]. In OLM, Schneider and others tie together situational emotions and learning, based on the flow theory; see Csikszentmihalyi in Ref. [22]. They describe situational engagement as the interplay of three situational emotions: feeling highly interested, highly challenged, and highly skilled to the task at hand; hypothesizing that the moment of situational engagement would be optimal for learning. Within that moment, they depict the relationships of nine other emotions, which they categorize to enhance (feeling active, successful, happy, enjoyment, and confident), detract from (feeling bored, and confused), or accelerate learning (feeling stress and anxiety) (Fig. 1).

Students' different academic emotions and engagement could have reciprocal and cyclic connections [5]. Figure 1 shows how in OLM two-way interactions between enhancers or accelerants, and optimal learning moments are deemed possible, whereas situational engagement is said to have a one-way connection towards and detractors away from optimal learning moments [22]. Situational engagement is the trigger for optimal learning moments, which are then expected to feed and feed off on learning enhancing and accelerating emotions. Schneider *et al.* [22] take learning detractors to be different than enhancers and accelerants in that they do not believe detractors would lead to optimal learning moments, thus the one-way

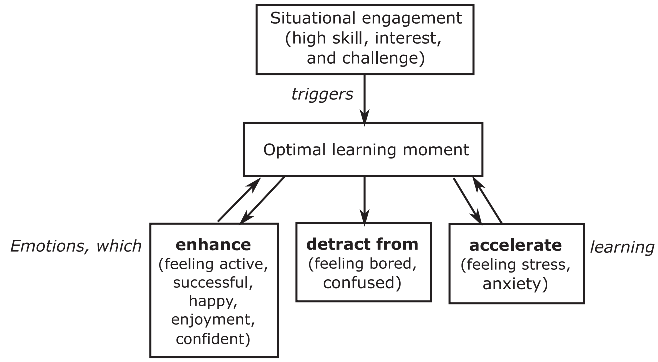

FIG. 1. Relationships between situational engagement, optimal learning moment, and other situational emotions according to the framework of optimal learning moment. Figure adapted after Schneider *et al.* [22] (Fig. 1).

relationship. OLM seems to define clearly the state of engagement, its antecedents, and consequences, which is important for a model of engagement [4].

Research using this model or part of it in high school context has found, e.g., regarding situational engagement: that of its components, challenge was connected negatively to enhancers and positively to detractors and accelerants, where as with skill and interest the connections were exactly opposite [22]; that certain science class room activities, such as developing models, constructing explanations, and reducing listening lectures can enhance the frequency of situational engagement [40,41]; but also that in the massive open online course environment, the sections where a teacher was explaining concepts were most engaging, perhaps because of their demand on deep learning strategies [42]. Regarding student retention, of all the twelve OLM emotions, only situational stress has been connected to leaving the physics track in high school [43]. However, when situationally engaged, students have been more likely to count the task as meaningful to themselves and their future [22], which in the long run, one would assume, would endorse student retention. As far as we know, only one study has applied OLM in situations both in and outside the classroom, with the results that co-occurrence of OLM emotions was consistent across situations [44]. Also, we found only one study where OLM was applied in the HE context [45]. Even though Schneider *et al.* [22] concentrate on high school students, they do not express that the utility of OLM would apply only in the secondary level education. The single study from the HE context found that situational engagement measured in physics tutorial sessions correlated with course exam results [45]. Hendolin [45] also tested the utility of the OLM concept in HE physics education by comparing their OLM measurements against standardized tests, and found a significant overlap in OLM frequency and how interested in physics students were, and how valuable physics was to them (personal application and relation to real world subcategory of CLASS [46]). He denotes the need for further studies concerning the utility of OLM measurements.

Schneider *et al.* [22] suggest a theoretical connection from OLM to science learning. In our study, we rely on the OLM model of situational emotions and continue the line of thought to more far reaching course passing, course grades, and student retention, within HE physics. Thus, our study widens the use of OLM in new areas and brings new perspectives on HE physics student retention.

C. Finnish HE system and student retention

In 2018, 4.2% of all new entrants in Finnish HE began their studies in natural sciences, mathematics or statistics, and of all Finnish HE graduates (91% bachelors degree) 5% were from natural sciences, mathematics, or statistics [2,3]. On all fields of HE in Finland, by the beginning of the second year of studies 91% were still enrolled to bachelor's or an equivalent program, after the theoretical duration of the program 43% had graduated but 14% were no longer enrolled in tertiary education, and given three years more 70% had graduated from bachelor's or equivalent but 18% had not nor were they enrolled any more [1]. Thus, on a general level the Finnish numbers are a bit brighter than OECD average.

Between the years 2000 and 2020 in Finnish universities, 11 500 students (35% women) started their studies in physics or in general science education programs, which in some universities include physics [47]. It is common that physics HE students in Finland have the Finnish high school education, from which one usually graduates at the age of 19 or 20. On the national level, over 85% of the students who started in HE physics programs and more than 65% starting in HE general science programs between 2000 and 2020 were from the age group 15–24 years [47], most likely around 20 years. Still, comparing the Finnish bachelor's degrees accumulated from physics and general science education HE programs to the amount of new entries on these fields four years earlier between 2005 and 2020 [47], it can be deduced that only 37.4% have graduated on a national level. This indicates that in Finland there might be more severe retention issues in physics than in other fields of HE on a national level.

The Finnish HE system can roughly be divided into traditional research and teaching intensive universities and universities of applied sciences. Finnish universities abide the European HE system, with a three year bachelor's and a two year master's degree. In 2018, more than 90% of first-time entrants to Finnish HE enrolled to bachelor's degree programs [3]. HE in Finland is mainly free of charge, all HE students are qualified for a monthly student allowance regulated by the government, and they can apply for student apartments, the rent of which are usually below the local average. We expect these forms of financial support to alleviate the effect of different economical backgrounds of the students, which is reported to be in part responsible of withdrawal from HE [48]. However, there are restrictions on the student allowance which state, e.g., that a student should accumulate roughly 60 study credits per academic year (nine months) [49]. One study credit equals to 27 hours of work, abiding the European Credit Transfer and Accumulation System (ECTS). This demand could cause substantial stress to students who progress slower.

D. Aim and research questions

The aim of this article is twofold. First, we aim to produce explorative insights about the interplay of being in learning situations, situational engagement, situational emotions, gender, course success, and student retention. Second, it will be done with data and methods that acknowledge the situational nature of engagement and emotions, while still drawing conclusions on an individual level. The novelty of this study is in both the framework being used in a rare context and the methods being used in a progressive manner. Three research questions are formulated:

1. How is being in a learning situation, experiencing situational engagement, and gender connected with situational emotions?
2. How are individual level differences in emotions and gender connected with course success?
3. How are course success, individual level differences in emotions, and gender connected with retention?

1. Hypotheses

In light of the previous discussion, the variables in focus and the hypothesized connections are visualized in Fig. 2.

The model in Fig. 2, is based on the situational emotions and engagement as defined in OLM [22]. Regarding the first research question, we hypothesize that in learning situations the learning accelerating emotions would be higher than in other situations [28]. We expect that learning enhancing and detracting emotions in learning situations would differ of those in other situations, but leave open whether the effect will be positive or negative. Situational engagement is expected to raise the enhancing and accelerating, and lower the detracting emotions. It is also anticipated that females would report higher detracting and lower enhancing emotions especially in learning situations [21,22]. We test whether this can be spotted as a direct or a nested effect.

On the second research question, we hypothesize that higher enhancing emotions, lower detracting emotions, and moderate accelerating emotions would result in better course success (more passed courses, higher grades) [6,14]. The same would apply to the third research question, with the addition that higher course success would result in greater probability of retention [8,21]. We test the direct effect of gender to course success and retention, but leave the hypothesis open as the previous research about the effects of gender seem contradicting.

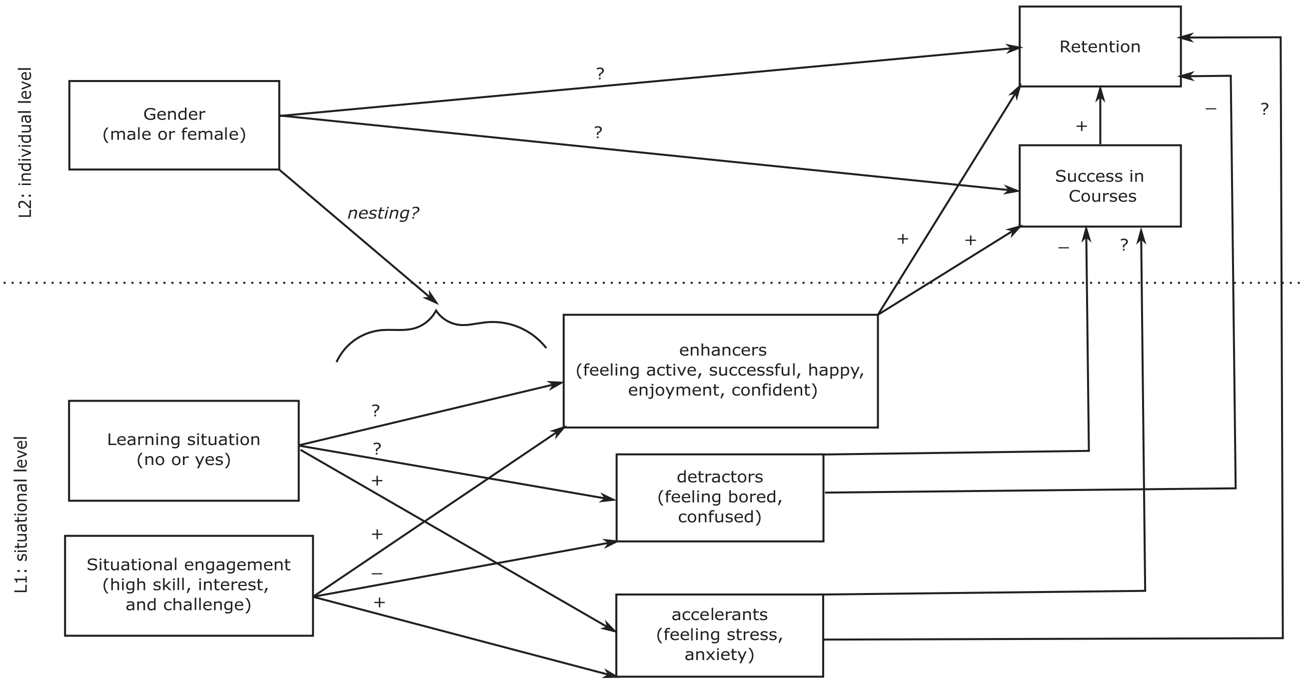

FIG. 2. Hypothesized connections of the chosen variables on two hierarchical levels. Plus sign (+) indicates an expectation of a positive effect and minus sign (−) a negative one. Question mark (?) indicates an expected but not further specified effect. Situational engagement and being in a learning situation are expected to affect situational emotions, and these effects might be nested within gender. The effects of situational emotions cross the line between situational (within-person, L1) and individual (between-person, L2) levels, and are expected to affect success in courses as well as retention. Gender could also have direct effects to emotions, success in courses, and retention.

III. METHODS

A. Context

The target university of this study is a public university, with both research and teaching being its main functions. In the university, the overall proportion of BSc degrees from any field compared to the amount of first year students four years earlier between 2005 and 2020 is 82.9%, the national level being 85.4%. At the same time, the BSc degrees from physics or general science programs compared to the amount of students starting four years earlier in these programs was 36.1%, which again very close to national level of 37.9%. The same retention problem is reflected in these numbers as in OECD reports earlier, with physics doing worse than many other subjects [47].

The gender distribution in the target university's physics or general science degree programs between 2000 and 2020 was the same as in the national level, about 35% women. In the same timeline, the new students in the target university were on average younger than on the national level, with 90%–94% of the new students belonging to age groups 15–19 or 20–24 [47].

The physics education program in the university has versatile orientation options for later advanced special studies, but the first year of studies we focus on consists mainly of common mandatory courses of physics and mathematics. Physics students have a choice between scientific or technical mathematics, as per recommendations of their physics orientation. Almost all courses accumulate 5 ECTS each and are divided throughout the year in five periods (two in fall, two in spring, and one in summer). The grading of courses is either pass or fail or a discrete scale of one to five, five being the highest grade. To graduate as BSc, students need to accumulate 180 ECTS. In the target programme, depending on the students physics orientation, they would need to pass nine ECTS worth of general studies such as language and orientation courses, at least 82 ECTS of physics, at least 40 ECTS of mathematics, at least 25 ECTS of physics courses directed at different physics orientations, and the rest are optional. If the student aims to become a teacher, 30 ECTS in the BSc study plan are replaced by 30 ECTS of pedagogical courses from the Faculty of Education. To graduate as MSc, the student continues to achieve 120 ECTS more, where at least 80 ECTS are advanced physics courses including a master's thesis (35 ECTS). In the teacher orientation, 60 ECTS need to be advanced physics courses including a master's thesis (20 ECTS) and pedagogical courses are continued by additional 30 ECTS. The rest of the 120 ECTS are optional major and minor subject courses, but there may be recommendations as per the student's orientation. It is recommended that the student accumulates 60 ECTS per year to be able to graduate in target time of five years (three years for

BSc and two years for MSc), but students have a law-based right to exceed this by two years [50], Sec. 41.

1. Participants

The 20 participants of this study (five women) belong to a single student cohort of the target university's physics education program. They were first year students and they participated to this study voluntarily, after the research was advertised in two occasions at the beginning of their studies. All in all 39 students commenced in this study, but the 20 final participants were chosen on the basis of producing enough data about their situational emotions, as is later explained in the data section. All participants signed a proof of consent, allowing us to use the data they produced during the data gathering period and their student records. 18 of the final 20 participants filled a background questionnaire revealing they were all in their early twenties with Finnish high school education. This is a very common scene in the target education program and we have no reason to assume the two participants not returning the background form would differ significantly in this respect.

The participants' prior preparation, physics skills, or study skills are not in the focus of this study, but it is acknowledged they could have an effect on experienced emotions and retention. All the participants had at least the minimum skills required to enter the target physics degree program, demonstrated either with taking part to an entrance examination measuring high school level physics skills, or with the matriculation exam results of physics being among the top 60% of the cohort. The entrance requirements may vary between universities and degree programs.

B. Measurements

The variables in Fig. 2 are divided in individual level (L2) and situational level variables (L1).

Gender (L2) is defined as a dichotomous variable (male = 0, female = 1). We acknowledge that polarizing the gender variable as such could obscure other identities, which might be included in the sample.

Course success (L2) is operationalized in two ways by using four mandatory courses, two from physics and two from mathematics, studied in the first fall semester. The physics courses, Physics Mathematics (Later: Phys1, 5 ECTS) and Mechanics Pt. 1 (Phys2, 3.5 ECTS) were mandatory to all participants. The students had a choice between scientific and technical mathematics. Two courses from each orientation were chosen: Introduction to Real Functions (Smath1, 5 ETCS) and Matrice Calculation (Smath2, 5 ECTS) from scientific math, and Basic Mathematics Pt. 1 (Tmath1, 5 ECTS) and Matrice Algebra (Tmath2, 5 ECTS) from technical math. These four courses were chosen by the authors to represent the core content of the first fall semester, without which the student might have difficulties proceeding. The first operationalization concerned only whether students had completed these courses or not (course failed = 0, course passed = 1). The second operationalization focused on the grades from these four courses (discrete 1–5 per course). If the student had not passed a course, the course grade was set to zero.

Retention (L2) is the final outcome variable. It is operationalized by whether the student stayed in their physics major during a three-year time frame or not, according to their student records (no = 0, yes = 1). The student would be categorized retentionwise as "no" if they had changed their major within or ceased to enroll as a student to the focus university. It would be beneficial to be able to separate transfer behavior from total drop out from HE, as it might invoke different practical implications [18]. As we acquired the data on retention from the university, we were not able to confirm whether those who quit physics continued to enroll in HE somewhere else, unless it was within the same university. For this reason, we use the more coarse dichotomous categorizing.

Learning (L1) is operationalized as the participant being in a learning situation or in some other situation according to their own reporting. In situational measurements, the participants were asked "What are you doing?" and given the following options to choose from: 1. learning, 2. at work, 3. at a hobby, 4. free time, 5. none of these. The responses were dichotomized so that choices 2–5. were combined as "not learning". Thus, learning is a dichotomous variable (not learning = 0, learning = 1).

Engagement (L1) and other situational emotions are operationalized using the OLM framework [22]. In situational measurements, students were asked to assess their emotions on a discrete scale from 1 (not at all) to 5 (very much). To be situationally engaged, student needed to report simultaneously high level of interest, high feeling of being challenged, and high feeling of having enough skills to perform [22]. High emotional level was determined as four or five. Engagement is a dichotomous variable (no situational engagement = 0, situational engagement = 1).

Other situational emotions are operationalized as three sum variables of nine distinct situational emotions [22]. *Enhancers* (L1) consist of feeling active, successful, happy, enjoyment, and confident. *Detractors* (L1) are the sum of feeling bored and confused. *Accelerants* (L1) combine the feelings of stress and anxiety. Again, all nine distinct emotions were measured on 1–5 discrete scale (1 = not at all … 5 = very much). The three sum variables were calculated by adding together the values of the respective distinct emotions and dividing this value with the number of the distinct emotions.

C. Data gathering

In data gathering we used experience sampling method [22,29] via the PACO mobile app installed to student's personal cellphone (www.pacoapp.com). The app sent prompts at random times between 8 am and 9 pm five times a week to the students to answer a short query about their emotions and if they were in a learning situation or not (Appendix A). Once receiving a prompt, the student had 15 min to answer, otherwise the prompt was discarded. This made sure the responses were as randomly situational as possible. The query took under a minute to answer.

Other data includes background questionnaire filled by the participants at the beginning of this study and student records received later on from the university, with the consent of the students. The participants were given pseudonymes for us to be able to connect different datasets, which will be eventually anonymized completely.

D. Data

The dataset consists of $N = 440$ responses from 20 students to the ESM questionnaire, their student records of three first years of studying ($N = 20$), and background information forms ($N = 18$). ESM is a demanding data collection method, since the participants need to commit to it for a longer period. Any longitudinal data collection is acclined to participant attrition. Our eight week data collection period was quite ambitious, as a typical ESM study might collect data from one to three weeks [29]. By the end of the second month of our data collection, more than 40% of our 20 participants were still responding to the prompts. ESM paired with hierarchical analysis can accommodate for unanswered prompts to a degree [29], and using these methods allowed us to keep more participants than just those who answered nearly all prompts, obviously with the cost of lower statistical power. Fisher's exact test [51] was used to ensure there were balanced amount of responses using gender, retention, learning, and engagement as categorizing variables.

1. ESM data: Situational emotions

The ESM responses are from the first two months of the students' HE physics studies. All in all 861 prompts were sent to 39 students and 520 responses received from 36 students. 123 of the responses did not have a scheduled timestamp. It may be that students responded after the timer had exceeded the 15 min, but there was still an indicator in their phones showing a prompt had been launched. The unprompted responses were included in the data as is. The amount of prompts a student had received during the two months varied between 0 and 43 ($M = 24$), as the choice of using the application and registering to the query or removing it was in their own hands. The amount of responses given by students, prompted or unprompted, varied between 0 and 39 ($M = 15$).

To be able to compare learning situations to other situations, the participants had to have responded to the questionnaire in both situations. It was decided that the participant had to have responded at least four times in a learning situation and at least four times in another situation. This decision narrowed the participants to the 20 students mentioned earlier.

These 20 students together received 668 prompts and left 451 responses (response rate 67.5%). Empty or nonresponses (217), and nine incomplete responses with one or two missing emotion values were discarded, leaving 442 complete ESM responses. Within these 442 responses, the amount of responses per student varied from 10 to 38 ($M = 22$, $\mathrm{SD} = 8.8$). 168 of these responses were from learning situations (38%) and 274 from other situations, such as working or leisure time. The students were further asked in the query what kind of learning situation it was. Out of the 168 responses, 34 (20.4%) were from physics lecture, 6 (3.6%) were from an instructional physics exercise event, 47 (28.1%) were from doing physics exercises outside the instructional events, 10 (6.0%) were sent while learning physics from literature or other material, nine responses (5.4%) were from learning physics in some other way, 54 responses (32.3%) were from learning some other subject than physics, and seven responses (4.2%) were from learning situations where none of the given options applied. No one responded "doing physics laboratory work," although that was one of the options. So, within the learning situations data there were clearly some responses from official, instructional settings (e.g., lecture) but also some that were from outside official class settings (e.g., doing physics exercises without instruction). Although it would be an interesting question to ask how different kinds of learning situations reflect in emotions, in this study, all learning situations are bundled together to avoid spreading the data too thin.

Alpha analysis [52,53] was used to test whether the grouping of emotions as sum variables was reliable. Enhancers ($\alpha = 0.800$) and accelerants ($\alpha = 0.849$) were fine. Detractors had a lower value of alpha ($\alpha = 0.506$) but considering the interitem correlation (0.339) and the fact that the group had only two items, detractors too were kept as a group [54] (p. 97). Later in the modeling phase, two ESM responses were discarded as extreme residual outliers; thus the main analysis includes $N_1 = 440$ situational responses from $N_2 = 20$ students.

The sum variable enhancers had measurements in 19 out of 21 possible categories, detractors and accelerants had measurements in nine out of nine possible categories. The distribution of enhancers was close to symmetric (skewness −0.529; kurtosis 0.591), as was that of detractors (0.540; −0.247). Distribution of accelerants was not symmetric (0.816; −0.240). The accelerant values were transformed to LOG10 values, which improved the symmetricity (0.098; −1.070). Thus, these three sum variables were considered as continuous instead of categorical [55] (p. 130).

2. Study records: Course success and retention

Of the 20 participants, 13 (65%) were continuing their physics studies by their third year. Seven students (35%)

TABLE I. Frequencies of course grades from Phys1, Math1, and Math2. Courses are graded in a discrete scale from 1 (lowest) to 5 (highest). No grade means the number of students, whose study records did not indicate passing the course.

	No grade	1	2	3	4	5
Phys1	5	1	0	4	4	6
Math1	4	2	0	1	5	8
Math2	5	1	2	2	4	6

had discontinued, six of them after the first year, and the last one by their third year.

14 participants (70%) had passed all the four first semester courses in the focus and one or two participants passed three to none of the courses (mean $= 3.15$, SD $= 1.46$). To simplify matters, the math courses were juxtaposed as Math1 (Smath1 or Tmath1) and Math2 (Smath2 or Tmath2). Passing any one of the four courses correlated strongly and statistically significantly with passing the other three courses. Thus, course success measured as passing core courses was kept as a single variable courses, with discrete value 0–4 indicating the number of the passed courses.

The course grades ranged from 1 to 5 in each course, means were 3.93 for Phys1 (SD $= 1.16$), 4.00 for Phys2 (SD $= 1.28$), 4.06 for Math1 (SD $= 1.34$), and 3.80 for Math2 (SD $= 1.32$). Phys1 and Phys2 grades had strong correlation (Spearman $\rho = 0.53$, $p < 0.05$), and Phys2 and Math2 grades had very strong correlation ($\rho = 0.72$, $p < 0.01$). Thus, Phys2 was left out as a variable, since it would not give any new information. Course success measured as course grades was kept as three separate course grades, Phys1, Math1, and Math2. The frequencies of grades per course are presented in Table I.

From Table I we can seem that most of the students who passed the courses, got high grades but there were always four to five students who had not passed one or more courses.

3. Gender

The sample consisted of fifteen males (75%) and five females. The sample resembles both cohort and larger local population gender distributions. The imbalance of genders is typical for physics, as Sax and others [17] have reported.

4. Missing data

The nine incomplete ESM responses were missing one or two values of the situational emotions. Little's MCAR test [56] showed that these values were missing completely at random (MCAR) both on all data level ($\chi^2 = 91.670$, $df = 76$, $p = 0.106$) and on individual level ($p > 0.05$ for all). The incomplete responses were 2.5% of all data. On the individual level the proportion of incomplete answers was between 2.6% and 15.0%. An advanced way of handling the missingness would have been multilevel multiple imputation [27], but this was not an option in SPSS v. 26. Multiple imputation without the multilevel notion, when data is hierarchical, might introduce biased results and so it was decided to use list wise deletion, which works for multilevel data with the MCAR missingness in outcome variables [57].

Grade variables were missing values whenever a student had not passed the course. These missing values were replaced with zero to indicate a "no grade." Imputing a 1–5 value would not have been a good choice, since the missing values had the meaning of not passing the course at all. Using zero as a marker let us keep these students as part of the data.

E. Analysis

The data were analyzed using hierarchical linear modeling with Griffin's two-step approach (HLM) [58], which provides results to all three research questions sequentially. Step 1 consists of traditional HLM, well explained by, e.g., Hox *et al.* [55] and Snijders and Bosker [59]. In this step, one L1 variable is set as the outcome, and other L1 and L2 variables are set as predictors. HLM differentiates the L1 and L2 residuals, making it possible to detect changes on either level.

When the outcome variable is not on the lowest level of the data hierarchy, as is the case with our second and third research questions, the choices of modeling and statistical analysis are not very established [60]. According to Becker, Breusted, and Zuber [60], by using the two-step approach of Griffin [58], one can model the entire chain of mechanisms between macro and micro levels without committing a measurement error. The method does not provide exact measures of how the different predictors affect the final higher level outcome, but it provides evidence whether the process exists [58]. Also, as there are only a few predictor variables in the first step, interpreting the results of the second step should not be overly difficult.

In the second step of the analysis, individual level residuals, recorded from step 1, were assigned as new L2 predictor variables. Inherent from step 1, the possible effects of learning, engagement, and gender were controlled in each of these residuals. These new L2 predictors, along with the existing ones were then regressed on passing courses, course grades, and student retention using single level regression methods.

The centering of predictor variables follows the suggestions of Enders and Tofighi [61]. When examining the direct, nested, or cross effects of L1 predictors, learning and engagement were centered within cluster ($\mathrm{Learning}_{\mathrm{cwc}}$, $\mathrm{Engagement}_{\mathrm{cwc}}$) by subtracting each student's individual means from the value of situational learning or engagement. When examining the direct effect of the L2 variables, learning and engagement were centered around the grand mean ($\mathrm{Learning}_{\mathrm{cgm}}$, $\mathrm{Engagement}_{\mathrm{cgm}}$). Gender was left uncentered.

The analysis was conducted with SPSS's v26 linear mixed models in step 1. In Step 2 R [62] was used, specifically the packages logistf [63] and brglm2 [64].

1. Assumptions and sample sizes

HLM tests the linear relationships between the predictor and outcome variables [55]. A visual exploration of the data as well as correlation tables gave indications that such relationships would exist.

Another assumption of HLM is that there is no multicollinearity between explanatory variables [55]. This was examined with Spearman's correlation and using collinearity statistics of linear regression ($N = 442$). The bivariate correlations between gender and learning ($\rho = -0.046$, $p > 0.05$), and gender and engagement ($\rho = -0.039$, $p > 0.05$) were negligible and insignificant. The correlation between learning and engagement was statistically significant and small ($\rho = 0.113$, $p < 0.05$). Putting gender, learning, and engagement as predictors and L1 emotions one by one as the dependent variable, the collinearity statistics confirmed no collinearity (all tolerances > 0.95, all VIF < 1.1).

Lastly, HLM assumes residuals should exhibit multivariate normal distribution, nonlinearity, and homoscedasticity [55]. These were checked twice during the HLM analysis, first time from the empty model and second time from the final model with box plots, normality tests, scatter plots, and one way ANOVA. At this point, two situational responses were removed as extreme residual outliers, resulting in the final L1 sample size of $N = 440$. The normality of the residuals of enhancers and detractors on both levels varied, but were overall tolerable. As to the residuals of accelerants, there was noticeable heteroscedasticity, which was countered by transforming the values of accelerants to LOG10 values [55], p. 157, which improved the situation. ANOVA tests confirmed there were no significant differences between individuals, in respect of the variance of their situational residuals.

The final sample size on the situational level ($N_1 = 440$) is sufficient for HLM, but the sample size on the individual level ($N_2 = 20$) could ideally be higher [55]. A simulation study by McNeish and Stapleton [65] suggests that using restricted maximum likelihood estimator (REML) with Kenward-Roger adjustment should produce acceptable regression coefficient and variance component estimates, and confidence interval coverages with even less than 20 clusters. However, the power of such analysis may be low [65]. Hox, Moerbeek, and van de Schoot [55] (p. 214) refer to a study by Bell, Morgan, Schoenenberger, Kromrey, and Ferron [66], where using REML with Kenward-Roger adjustment for degrees of freedom produced only minimal bias in fixed effects coefficients, and the type I error rates and confidence interval coverages for fixed effects were slightly conservative. Thus, in the HLM analysis, the combination of REML with Kenward-Roger adjustment was used and it was expected that only moderate to strong effects could be found.

The assumptions of multinomial and binary logistic regression used in the second step of the analysis expect only that the outcome variable is mutually exclusively categorized and that there exists no multicollinearity between predictor variables. Correlation and VIF checks were made at relevant phases.

It has been suggested [67], that the Firth method, which introduces a penalizing term to ML estimation equation [68], should be preferred over ML in logistic regression, when the sample size is small. Bull *et al.* [69] expand the Firth method to cover multinomial logistic regression, with good results on reducing ML asymptotic bias with sample size as small as 25. In our analysis, to counter the possible bias of the regression coefficients due to sample size in step 2, multinomial regression with bias reduction by Jeffrey's prior was used [64,69].

2. Step 1: Hierarchical linear modeling

The first step started with an empty model M0, consisting of an intersection term β_{0j} and a L1 residual term e_{ij}, producing a level 1 regression equation for outcome X_{ij} for a student j in a situation i:

$$X_{ij} = \beta_{0j} + e_{ij}. \tag{1}$$

In Eq. (1) the term β_{0j} is the within individual mean of the situational variable X_{ij} for student j and the term e_{ij} is the situational level residual of that mean. To include between individuals aspect, the intersection term β_{0j} is divided into between individuals mean γ_{00} and its individual level residual u_{0j}:

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{2}$$

The empty model $M0$ was run separately for enhancers, detractors, and accelerants as the outcome variable X_{ij}. The empty model produces variances of u_{0j} and e_{ij}, which can be used in calculating the intraclass correlation coefficient (ICC) by dividing $\mathrm{var}(u_{0j})$ by the total variation $\mathrm{var}(u_{0j}) + \mathrm{var}(e_{ij})$. This indicates the proportional part of between individuals variation against total variation and it indicates whether the multilevel analysis is meaningful.

Next, the L1 predictors learning and engagement, and the L2 predictor gender were added one by one to the model. Then, the hypothesized interaction terms of learning nested within gender and engagement nesting within gender were added. The slope variances of the L1 predictors were tested, as were the aggregated L1 variables as L2 predictors (individual mean of learning, individual mean of engagement), as suggested by Snijders and Bosker [59]. They were interpreted as the proportion of the measurement situations the participant reported learning or engagement.

In step 1 the values of these two variables were simply dichotomized above and below sample average (below = 0, exact or above = 1), to describe whether the student was found more or less situationally engaged and in learning situations proportionally, than the sample mean. The effect of small covariance of engagement and learning, spotted in preliminary analysis, was tested using an interaction term engagement by learning. All this was modeled three times, placing enhancers, detractors, and accelerants as the outcome variable one at a time.

The following is an example of a fuller model tested:

$$X_{ij} = \beta_{0j} + \beta_{1j}(\mathrm{Learning})_{ij} + \beta_{2j}(\mathrm{Engagement})_{ij} + e_{ij}, \tag{3}$$

where

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{Gender}_j + u_{0j},$$
$$\beta_{1j} = \gamma_{11}\mathrm{Gender}_j + u_{1j},$$
$$\beta_{2j} = \gamma_{21}\mathrm{Gender}_j + u_{2j}.$$

In Equation (3) the term β_{0j} consists of the between individuals mean γ_{00} with the added coefficient γ_{01} to model the fixed effect of gender on X_{ij} and ending with the individual level residual u_{0j}. Term β_{1j} is the effect of L1 predictor variable learning, which can be divided to the fixed cross-level effect γ_{11} of learning being nested within gender, and the slope variance u_{1j}. Lastly, there is the term β_{2j}, consisting of fixed cross-level effect of engagement nested within gender γ_{21} and its slope variance u_{2j}.

3. Step 2: Logistic regressions

Along with answering the first research question, step 1 of the analysis produced three new individual level variables: the between individuals residuals $u_{0j,\mathrm{enh}}$, $u_{0j,\mathrm{det}}$, and $u_{0j,\mathrm{acc}}$ (later u_{enh}, u_{det}, and u_{acc}) from the best fitting models. The effect of any of these u_{0j} terms on the final L2 outcomes can be interpreted as the aggregated net effect of the L1 variable X_{1ij} chosen as the outcome in the first step, and all other L1 and L2 covariates $X_{2ij}, \ldots, X_{nij}$ and Z_j [60]. As Griffin [58] puts it, the residual parameters u_{0j} describe the cluster level properties based on lower level measurements. In our study, the contextuated interpretation is that the L2 residuals (u_{enh}, u_{det}, and u_{acc}) describe the individual level emotions compared to the sample average (the intercept), when gender, engagement, and learning have been controlled. In other words, they are the individual adjusted levels of the emotions, truly excluding the situational fluctuation.

Following Griffin's [58] logic, the three L2 residuals were used as new L2 predictor variables alongside previously existing L2 predictors (gender, mean of learning, mean of engagement) in a single level multinomial logistic regression. The outcome variables in step 2 were courses (the number of passed courses out of the four), Phys1 grade, Math1 grade, Math2 grade, and retention. Since producing new predictors, the assumption of no multicollinearity was checked again and it was found that u_{det} correlated strongly and significantly with both u_{enh} (Pearson's $r = -0.66$, $p < 0.01$) and u_{acc} (Pearson's $r = 0.69$, $p < 0.001$). Thus, only u_{enh} and u_{acc} were added to the list of L2 predictors. Three example equations are presented below [Eqs. (4)–(6)]. The symbols in the following equations are kept the same as in step 1 individual level terms to ease connecting of these equations to individual level phenomena.

$$\mathrm{logit(Courses)} = \gamma_0 + u_{0j}. \tag{4}$$

Modeling for any outcome began with an empty model, consisting of only the intercept γ_0 and the individual level variance u_{0j} [Eq. (4)].

$$\mathrm{logit(Courses)} = \gamma_0 + \gamma_1\mathrm{Gender} + \gamma_2 u_{\mathrm{enh}} + \gamma_4 u_{\mathrm{acc}} + u_{0j}. \tag{5}$$

When courses or any of the course grades was set as the outcome, different models were tested where gender, mean of learning, mean of engagement, and the individual level residuals u_{enh} and u_{acc} were set as predictors one by one [e.g., Eq. (5)].

The regressions concerning grades were run twice, returning the zeros to NAs in the second round. This was done to assure that the special meaning of zero linked to passing or failing the course was not blurring the results with the cost of lowering N from 20 to 15.

$$\mathrm{logit(Retention)} = \gamma_0 + \gamma_1\mathrm{Gender} + \gamma_2 u_{\mathrm{enh}} + \gamma_4 u_{\mathrm{acc}} + \gamma_5 Courses + e_j. \tag{6}$$

When retention was set as the outcome variable, all the same predictors as with the previous outcomes (courses and grades) were tested as predictors, but furthermore, the variables courses and grades were tested as additional predictors for retention [e.g., Eq. (6)].

Finally, a note must be made about two of the predictors: since mean of engagement meant the proportion of responses in which the student reported situational engagement, and mean of learning meant the proportion of responses the student reported they were in a learning situation, these predictors are inherently bound between 0 and 1. Regarding the interpretation of the logistic regressions, an increase of 1 when talking about either of these means would mean an increase of 100% units, which is not realistic. These variables were transformed to 10* mean of engagement and 10* mean of learning, which describe the same proportions as before, but now the values could vary between 0 and 10. Thus, an increase of 1 in either of these predictors in a model would mean a 10% unit increase in the proportion of learning situations or situational engagement, which is a much more realistic scenario. It was checked that transforming the predictor variable values this way did not change the model indicators.

4. Strategies for choosing the best fitting models

Since using the REML estimation, the significance of the predictors and the model fit were assessed using the p values of the t tests for the fixed part of the model and chi-square tests on the differences in deviance statistics (−2LL) for the random part [55,59]. Any statistically significant predictors were left in the best fitting model. Any predictors causing the model not to converge (even after trying to adjust the number of iterations) were left out. When testing the nesting effect within gender, it was left to the best fitting model only if the effect on both genders was statistically significant. Effect sizes of fixed effects were estimated with the acknowledgement of the multilevel structure of data and thus of variances on two separate levels [70].

In step 2 of the analysis, the Akaike information criteria (AIC) was used to find the best model so that the lowest AIC value would indicate the best model fit, with the following additions: if the difference between two models' AICs (ΔAIC) was less than 2, there would be substantial support for both; if ΔAIC was between 4 and 7, there would be only little support for the model with the higher AIC; and if ΔAIC were over 10, there would be no support for the model with the higher AIC [71].

IV. RESULTS

A. Engagement and gender-nested learning affect situational emotions

The first research question asked, *how learning situation, situational engagement, and gender are connected with situational enhancers, detractors, and accelerants*. Common results were that individual level predictors gender, mean of learning, and mean of engagement did not have a statistically significant direct effect to any of the three situational emotions, nor did the cross effect of learning and engagement. However, situational level predictors learning and engagement did separately affect the situational emotions, and the effect of learning was nested within gender. Next, three models are presented for each of the three situational emotions explaining the mentioned significant results in detail. First is an empty model (M0) consisting of the intercept, within-individual, and between-individual variances. The second model (M1) is a best fitting model, when using only L1 variables (engagement and learning). The third model (M2) is a best fitting model when including gender as a nesting variable. Tables of all the models tested, including the results on L2 direct effects and L1 cross effects, can be found in Appendix B.

1. Enhancers

Being in a learning situation and being engaged had separate direct effects on the enhancers, and the effect of the learning situation was nested within gender, as presented in Table II.

From Table II we can see that for enhancers, the proportion of variance between individuals [$\mathrm{var}(u_{0j})$] was 38% of the total variance, indicating that hierarchical modeling was a meaningful choice ($M0$). Being in a learning situation lowered the score of enhancers by −0.46 points ($M1$). Learning was nested within gender, showing that the fixed effect of learning was about 0.22 points greater for the females ($M2$). The slope variance of learning was very small [$\mathrm{var}(u_{1j}) = 0.050$] and statistically significant, meaning that the degree of the effect of learning

TABLE II. Enhancers as the outcome for three different models. The values presented in the table are all statistically significant ($p < 0.05$), the significance of slope variance was tested with chi-square test using −2LL.

		$M0^a$		$M1^b$		$M2^c$	
Fixed		Coef	(SE)	Coef	(SE)	Coef	(SE)
Intercept	γ_{00}	3.531	(0.106)	3.530	(0.106)	3.530	(0.106)
$\mathrm{Learning}_{\mathrm{cwc}}$	γ_{10}			−0.461	(0.075)		
$\mathrm{Engagement}_{\mathrm{cwc}}$	γ_{20}			0.329	(0.087)	0.333	(0.087)
$\mathrm{Learning}_{\mathrm{cwc}}$ (Gender)	γ_{11}					−0.409 (m)	(0.085)
						−0.630 (f)	(0.152)
Random							
L2 variance	$\mathrm{var}(u_{0j})$	0.207	(0.073)	0.210	(0.073)	0.210	(0.073)
L1 variance	$\mathrm{var}(e_{ij})$	0.340	(0.023)	0.276	(0.020)	0.276	(0.019)
Slope variance of $\mathrm{Learning}_{\mathrm{cwc}}$	$\mathrm{var}(u_{1j})$			0.050	(0.034)	0.047	(0.034)
ICC		37.8%					
−2LL		827.259		756.978		757.028	

[a] $\mathrm{Enh}_{ij} = \mathrm{intercept} + u_{0j} + e_{ij}$.
[b] $\mathrm{Enh}_{ij} = \mathrm{intercept} + \mathrm{Learning}_{\mathrm{cwc}} + \mathrm{Engagement}_{\mathrm{cwc}} + u_{0j} + u_{1j} + e_{ij}$.
[c] $\mathrm{Enh}_{ij} = \mathrm{intercept} + \mathrm{Learning}_{\mathrm{cwc}}(\mathrm{Gender}) + \mathrm{Engagement}_{\mathrm{cwc}} + u_{0j} + u_{1j} + e_{ij}$.

TABLE III. Detractors as the outcome for three different models. The values are all statistically significant ($p < 0.05$), the significance of slope variance was tested with chi-square test using −2LL.

		$M0$[a]		$M1$[b]		$M2$[c]	
Fixed		Coef	(SE)	Coef	(SE)	Coef	(SE)
Intercept	γ_{00}	3.531	(0.106)	2.139	(0.106)	2.139	(0.106)
$\text{Learning}_{\text{cwc}}$	γ_{10}			0.583	(0.094)		
$\text{Engagement}_{\text{cwc}}$	γ_{20}			−0.303	(0.108)	−0.303	(0.109)
$\text{Learning}_{\text{cwc}}$ (Gender)	γ_{11}					0.569 (m)	(0.110)
						0.628 (f)	(0.196)
Random							
L2 variance	$\text{var}(u_{0j})$	0.201	(0.072)	0.204	(0.072)	0.204	(0.072)
L1 variance	$\text{var}(e_{ij})$	0.519	(0.036)	0.424	(0.030)	0.424	(0.030)
Slope variance of $\text{learning}_{\text{cwc}}$	$\text{var}(u_{1j})$			0.080	(0.057)	0.088	(0.061)
ICC		27.9%					
−2LL		1005.86		936.82		937.93	

[a]$\text{Det}_{ij} = \text{intercept} + u_{0j} + e_{ij}$.
[b]$\text{Det}_{ij} = \text{intercept} + \text{Learning}_{\text{cwc}} + \text{Engagement}_{\text{cwc}} + u_{0j} + u_{1j} + e_{ij}$.
[c]$\text{Det}_{ij} = \text{intercept} + \text{Learning}_{\text{cwc}}(\text{Gender}) + \text{Engagement}_{\text{cwc}} + u_{0j} + u_{1j} + e_{ij}$.

was slightly varying between individuals. Experiencing engagement increased enhancers by 0.33 points ($M1$).

2. Detractors

Being in a learning situation and being engaged had separate effects on the detractors too, and again the effect of learning was nested within gender. Results for detractors as the outcome are in Table III.

From Table III it can be deduced that for detractors, the proportion of variance between groups versus total variance was 28% and so the hierarchical modeling was justified ($M0$). Being in a learning situation increased detractors by about 0.58 points ($M1$). Learning was again nested within gender but showed that the difference was only 0.06, which means basically the same effect for both genders ($M2$). Again the slope variance of learning [$\text{var}(u_{ij})$] was small and statistically significant, indicating only a slight variation in the degree of the effect of being in a learning situation between individuals. Being engaged affected detractors negatively, by −0.30 points.

3. Accelerants

As explained in the data and analysis sections, the values of accelerants were transformed to LOG10 values to reduce their right skew and heteroscedasticity of situational residuals e_{ij}. The analysis process did not otherwise differ for them, but the results, which are presented in Table IV, are interpreted a bit differently than those of enhancers and detractors. Again, learning was found to be a statistically significant predictor and nesting within gender, but this time engagement was not part of the best fitting models.

TABLE IV. LOG10 values of accelerants as the outcome for three different models. The values are all statistically significant ($p < 0.05$), the significance of slope variance was tested with chi-square test using −2LL.

		$M0$[a]			$M1$[b]			$M2$[c]		
Fixed		Coef	(SE)	10^{coef}	Coef	(SE)	10^{coef}	Coef	(SE)	10^{coef}
Intercept	γ_{00}	0.313	(0.033)	2.055	0.313	(0.033)	2.056	0.313	(0.033)	2.056
$\text{Learning}_{\text{cwc}}$	γ_{10}				0.108	(0.016)	1.282			
$\text{Learning}_{\text{cwc}}$ (Gender)	γ_{11}							0.103 (m)	(0.018)	1.267
								0.126 (f)	(0.033)	1.337
Random										
L2 variance	$\text{var}(u_{0j})$	0.020	(0.007)	1.047	0.020	(0.007)	1.047	0.020	(0.007)	1.047
L1 variance	$\text{var}(e_{ij})$	0.026	(0.002)	1.061	0.023	(0.002)	1.055	0.023	(0.002)	1.055
ICC		43.8%								
−2LL		−302.801			−341.619			−341.619		

[a]$\text{Acc}_{ij} = \text{intercept} + u_{0j} + e_{ij}$.
[b]$\text{Acc}_{ij} = \text{intercept} + \text{Learning}_{\text{cwc}} + u_{0j} + e_{ij}$.
[c]$\text{Acc}_{ij} = \text{intercept} + \text{Learning}_{\text{cwc}}(\text{Gender}) + u_{0j} + e_{ij}$.

TABLE V. Results of multinomial logistic regressions for models with substantial support, when the outcome was the number of passed courses ($N = 20$, compared to courses = 0).

Model	*Courses*	*Coef*	OR = e^{coef}	AIC	*Residual deviance*
Intercept + u_{0j}				128.7	40.7
Intercept + $u_{\text{enh}} + u_{0j}$	1	−0.14	0.87	130.5	34.5
	2	−0.03	0.97		
	3	4.89	132.46		
	4	1.84	6.31		
Intercept + $u_{\text{acc}} + u_{0j}$	1	−2.63	0.07	130.5	34.5
	2	−3.25	0.04		
	3	9.58	14 472.21		
	4	−5.49	0.004		

In Table IV, the column 10^{coef} describes the proportional change of the accelerants caused by the respective predictor changing by 1. The proportion of variance attributed to differences between the group was about 44%, giving reason to apply HLM. Being in a learning situation caused an increase of 28% in the score of accelerants ($M1$). This would be in the magnitude of 0.58 points, calculated from the $M1$ intercept value. The fixed effect of learning was nested within gender, showing that the effect was seven percentage units greater for females ($M2$). In points, it would mean roughly 0.14 points higher reporting from females.

4. Explained variances and effect sizes

Adding the fixed L1 predictors (learning and engagement) to models affected mainly the random situational level variance. In calculating the variance explained and total effect size in all M1s, both level variances were considered to be bias free [70]. For enhancers the total variance explained was $R^2 = 0.11$ and effect size $f^2 = 0.13$. For detractors the explained variance was $R^2 = 0.13$ and effect size $f^2 = 0.15$. For accelerants the explained variances were $R^2 = 0.07$ and effect size $f^2 = 0.07$. Thus, the total effect size was medium for enhancers and detractors, and between small and medium for accelerants [70,72].

B. Emotions, mean of engagement, and gender predict success in courses

The second research question was *how are individual level differences in emotions and gender connected with course success*. The first set of multinomial logistic regressions targeting this question included gender, mean of engagement, mean of learning, and the two individual level emotion residuals u_{enh} and u_{acc} as predictors, and the number of passed courses as the outcome. None of the models tested were clearly better than the empty model, however, there was substantial support for models with either u_{enh} or u_{acc} as the predictor. Results from these models are presented in Table V.

From Table V we can see that having an increase of 1 in individual level enhancers would not make a difference between passing none, one, or two courses as the odds ratio (OR) is close to 1. However, higher enhancers would make it much more probable to pass three or four courses, compared to passing none (OR highly positive). Regarding accelerants as the predictor, having higher accelerants would make passing one, two, or four courses instead of none highly improbable (OR close to 0). But, comparing passing three courses against none, the higher accelerants would predict the first event (very high OR).

The second set of regressions concerning the second research question had the same predictors as previously, and the Phys1, Math1, and Math2 grades one by one as the outcome. For each, empty models had again the lowest AICs and thus the best fit, regardless of the grade scale used (0–5 or 1–5, zero indicating not passing the course). For Phys1 as the outcome, the second best model had gender as a predictor along the intercept and the residual (ΔAIC = 2.3, scale 1–5), however the results are omitted as not substantially supported.

For the outcome Math1, using the scale 1–5 brought up two models, substantially supported by the data, to accompany the empty model. The results from these models are presented in Table VI.

TABLE VI. Results of multinomial logistic regressions for models with substantial support, when the outcome was grade of Math1 ($N = 16$, compared to Grade = 1).

Model	Grade[a]	*Coef*	OR = e^{coef}	AIC	*Residual deviance*
Empty model				144.7	36.7
Gender	3	1.10	3.00	148.1	32.1
	4	−2.40	0.09		
	5	−0.96	0.38		
u_{enh}	3	2.44	11.51	147.3	31.6
	4	2.87	17.57		
	5	3.14	23.21		

[a]None of the participants got Grade = 2 in this course.

TABLE VII. Results of multinomial logistic regressions for models with substantial support, when the outcome was grade of Math2.

	Compared to Grade = 0, $N = 20$				Compared to Grade = 1, $N = 15$			
Predictor	Grade	Coef	OR = e^{coef}	Res dev	Grade	Coef	OR = e^{coef}	Res dev
Empty model				65.7				43.2
10 * Mean of engagement	1	1.50	4.49	57.6				
	2	0.55	1.74					
	3	0.54	1.72					
	4	0.86	2.37					
	5	0.41	1.50					
10 * Mean of learning	1	0.50	1.65	56.9				
	2	−0.87	0.42		2	−1.01	0.36	36.6
	3	−0.96	0.39		3	−1.08	0.34	
	4	0.44	1.55		4	−0.01	0.99	
	5	−0.30	0.74		5	−0.55	0.58	

From Table VI we find that being female made it 3 times more likely to get a grade of 3 instead of grade of 1. The effect was opposite for higher grades: being female made it 90% less likely to achieve a grade of 4 and 60% less likely to get a grade of 5. However, the effect of gender is not nearly as strong as that of individual enhancers: compared to achieving a grade of 1, an increase of 1 in enhancers would make it from over 10 to over 20 times more likely to achieve a grade of 3–5, with increasing odds.

For the course Math2 with 0–5 scale, there were two models which had substantial supportive evidence from data alongside the empty model. When the grade scale was chosen as 1–5, only one model besides the empty one got substantial support. The results of these models are presented in Table VII.

The results in Table VII tell that an increase of 10% units in the situations where the student reported situational engagement, would mean that they would get a grade of 1–5 1.5–4.5 times more likely than not to pass the course (grade = 0). The biggest difference was with getting the lowest grade against no grade.

An increase of 10% unit in situations the student reported they were learning, brought mixed results on the likelihood of getting any grade against not passing the course. The likelihood of getting a grade of 1 or 4 is higher, but lower in other cases. On the condition that the student had passed the course and had a grade of 1, the probabilities were against getting a 2, 3, or 5, but even in getting a 4 (Table VII).

C. Enhancing emotions and physics grades predict retention

The third research question asked, *how are course success, individual level differences in emotions, and gender connected with retention*. In short, the grade in the physics course and the individual level enhancing emotions both affected retention positively. However, the OR of staying against leaving physics when enhancing emotions were higher was more than double compared to the OR when the grade of physics was higher. Next are the details of these results.

In this last set of regressions, retention was set as the outcome with gender, mean of engagement, mean of learning, the two emotion residuals, courses, and the three course grades as predictors. Two models, with either the residuals of enhancers or Phys1 grade as the predictor were statistically significantly better than the empty model, according to profile penalized log-likelihood tests ($p < 0.05$). The results for the Firth logistic regression with retention as the outcome are summarized in Table VIII.

Table VIII shows that the difference of 1 in the residuals u_{enh} made it 12 times more likely that the participant had continued studies than quit. Having a one step better grade in Phys1, when measured on a 0–5 scale, made it 2 times more likely that the participant continued. Leaving out the zero category from Phys1 revealed a stronger effect: on the condition that the participant had passed the course and thus had a grade of 1–5, a step up in grade scale made it 5 times more likely to continue than quit. There were participants, whose residuals u_{enh} had a difference of 1, or even more, and the grades of Phys1 ranged from not passing the course to highest grade categories, making these results meaningful among our participants.

TABLE VIII. Comparison of logistic regressions for statistically significant predictors ($p < 0.05$) the individual level enhancing emotions (u_{enh}), and the grade of Phys1 in two different scales, with retention as the outcome. Odds ratio for enhancing emotions as the predictor is the highest. The intercept values are omitted as irrelevant.

Predictor	N	Coef	(SE)	OR = e^{coef}
u_{enh}	20	2.493	(1.378)	12.1
Phys1 (0–5)	20	0.645	(0.298)	1.9
Phys1 (1–5)	15	1.666	(1.024)	5.3

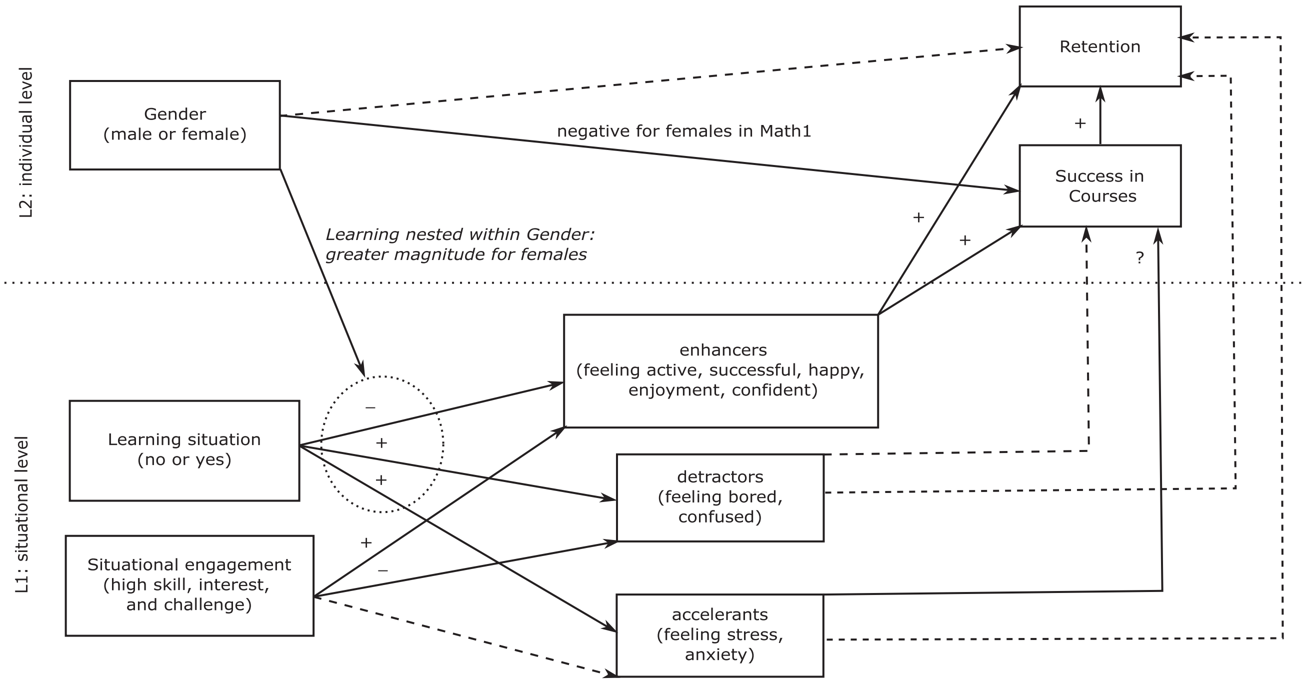

FIG. 3. Summary of results. Data supported connections are marked with solid arrows and unsupported hypotheses with dashed arrows. Plus sign (+) indicates a positive connection, minus sign (−) a negative one. Question mark (?) indicates a more complicated connection.

A summary of all the results can be found in Fig. 3, where the arrows present the tested connections and elaborating notes accompany them. Solid lines have been transformed into dashed ones, if the data did not show any support.

D. Limitations

The results presented describe the participating sample and are not to be generalized as is to wider populations or different types of focus groups. Limiting the participants by the amount of responses they gave to the ESM query may have biased the sample to be more homogenous than the population. The students' math orientation could have been taken as another individual level variable using the categories technical or scientific orientation, since it is possible that students who plan to leave physics, might choose technical orientation more often, as it could be more relevant choice in many majors close to physics, such as engineering. This question was out of the scope of this study, as it did not concentrate on study choices as the predictor for course success or retention. The standard errors of the coefficients in step 1 are quite high in some cases, probably as a result of the N_2 sample size. In the documentation of the R package brmultinom [64] there is a mention of the method being a "proof of concept," which could introduce unknown bias to the results. In both steps, there were several predictor variables, whose effect remained unclear, which is not to be interpreted as no effect at all, but as the need for further studies. We acknowledge the restrictions caused by the L2 sample size but remain confident that by choosing HLM as the analysis method we gained enough power to explore the rich multilevel connections between situational and individual predictors, and situational and individual level outcomes. Thus, the significant results are seen as indicators of the existence of the respective hypothesized connections, verified by the data.

V. DISCUSSION

The first research question was whether being in a learning situation, being situationally engaged, and the students' gender would be connected to situational emotions. Individual level variables (gender and the proportion of situational engagement or learning situations) were not statistically significant direct predictors for situational level emotions (Appendix B: Tables IX–XII). From a technical perspective, McNeish and Stapleton [65] mention how a low number of higher level groups in hierarchical modeling produces the soundest results on the lowest level. A similar study with more students could bring more light to this issue.

Being in a learning situation was connected to all three situational level emotions: to learning enhancers negatively, to learning detractors and accelerants positively (Tables II–IV). This result confirmed the hypothesis that learning accelerating emotions (stress, anxiety) would get higher ratings in learning situations, and that the enhacers (feeling active, successful, happy, enjoyment, and confident) and detractors (feeling bored and confused) would be affected. This was a within person result, meaning that these significant changes of emotions happened across the participants compared to their own average levels. Previous research has connected several positive and negative emotions to, e.g., self-regulated learning [6,15]. Our results confirm how a learning situation in itself can arouse a variety of emotions. It is important for the instructors to be aware that these kind of emotions can exist in a learning environment, whether the students show them openly or not.

Although gender was not a significant direct predictor for emotions, the effects of learning situation were nested within gender. Being in a learning situation affected the situational emotions to a different degree for women and men (Tables II–IV). For women, the effect of the learning situation was more negative on enhancers and more positive on detractors and accelerants, compared to men. Gender bound cultural differences might emerge in women reporting more extreme emotions than men [23]. However, by using the hierarchical analysis our results reflect the pure situational effects, even if women would have responded in a wider scale compared to men (but there was no direct effect). Other cultural features could play a role here too, since Schneider *et al.* [22] found that for high school students from the U.S. gender had a direct effect on emotions in a similar study, but for Finnish high school students it did not. As physics in general is a more male dominated field [17], the nested gender effect found could be a signal about students who differ from the "normal" and thus might have difficulties integrating socially and academically, and in developing identity as belonging to the discipline [73]. Learning environments could be particularly detrimental to underrepresentated students, such as women in physics [74], in agreement with our results. A recent study shows how, even when women were not the minority in a physics course, the sense of belonging was significantly increased in men and not in women by the end of the course [16]. Thus, gender and emotions in the context of physics HE education continues to be a relevant research subject, which could benefit of the perspective of more detailed local cultural factors and how they emerge in learning situations.

We did not restrict learning situations to mean the same as studying HE physics in classrooms. Relying on the randomness of our ESM data collection, it was assumed that the reported learning situations happened both in official, instructional classroom settings and in unofficial, uninstructional settings. This was confirmed by students reporting being in learning situations in different contexts such as in a lecture (official setting) and doing exercises unsupervised (unofficial setting). Previous research has indicated that in and outside school measurements in the high school level would produce a similar co-occurrence of emotions [44]. In our design, the particular co-occurrences of emotions stated above were connected to learning situations in general, regardless of whether these situations happened within official setting or elsewhere. Although we did not explore whether *where* the student was when in a learning situation affected the emotions, we suggest that, if emotions are considered an important element when designing teaching, the thought should cover the intended learning situations in and outside official settings.

According to the results, if situational engagement was reported (high interest, high skill, high challenge), learning enhancing emotions were reported more positive and detractors more negative, compared to measurements with no situational engagement (Tables II and III). These results describe the within person level, meaning that the change of emotions were compared to each participant's own averages and found significant. The change in the enhancers is in line with previous research showing co-occurrence of situational engagement and most of the measured enhancing emotions (enjoyment, success, and happiness) in high school [44], suggesting similarities across educational levels. The implications are that in designing HE physics teaching, pursuing situational engagement could be a way to affect situational learning enhancers and detractors. Detractors and behavioral engagement have been shown to connect negatively: boredom would lead to low behavioral engagement [5], and the opposite of engagement, disaffection [75]. An interesting question is could there be a chain of effects from situational affective engagement [22] through situational emotions (e.g., boredom) to resulting behavioral engagement?

The hypothesis of situational engagement's positive connection to accelerants was left unresolved (Appendix B: Table XI). The components of situational engagement (interest, skill, challenge) have been shown to separately have opposite connections to enhancers, detractors, and accelerants [22], which might have clouded our results too. The lack of evidence on the connection between situational engagement and accelerants could thus be caused by a more complicated connection than which was tested in this study. In the future, if situational engagement is defined as here, it might be insightful to keep its components apart, at least for comparison. Also, since stress has been mentioned as a driving force when in a moderate level [76], perhaps situational engagement does not raise accelerants, but evens out the extremities towards moderation?

Low correlation and nonsignificant cross effect of situational engagement and learning situation towards situational emotions (Appendix B: Tables IX–XII) implies that the effects of these predictor variables were truly independent of each other. This indicates that situational engagement and learning situations need to be considered as separate entities in designing teaching, if the goal is to affect situational emotions. In other words, there can be situational engagement outside learning situations and learning situations are not inherently engaging, but both can affect several emotions surrounding the process of learning. Ideally, situational engagement would overlap with learning situations.

So far we have discussed about the within-person results: the changes of emotions happening within a person compared to their own mean values. But why should we care about what their emotions were? The core of the second and the third research question was to find out if there were concrete consequences of having high or low levels of enhancers, detractors or accelerants, concerning course success or student retention. Because of the methods used, we are now crossing the level from within person to between persons.

The second research question asked if individual level emotions, gender, and the proportion of situational engagement or learning situations would be connected to course success, measured as passing a set of chosen four courses and their grades. The significant results were that higher enhancers might predict higher likelihood of passing more courses (Table V) and getting higher grades on a math course (Table VI). Positive emotions together with self-regulated learning and motivation have been proved before to be a significant factor for passing tests and having a higher GPA [6,15], as has sense of belonging [16]. Even though our model did not include the same mediating variables, the overall effect is in line with the mentioned research. The advancement is in measuring the emotions for a longer time period and excluding the situational variance. The effect of accelerants on passing courses was supported by the data (Table V) but it seems to be more complicated than a linear one. Anxiety has been connected earlier negatively to study outcomes, not directly, but mediated through planning, monitoring, and evaluating one's learning [14]. The nature of stress and anxiety have been theorized as beneficial to a point and harmful after that [76], implying perhaps a U-shape connection. Perhaps the research design would have benefited from defining the midrange values of these emotions as the preferred category and extreme values as one (extreme) or two (low or high) unpreferred categories, thus giving the measurements a different order than the linear 1 to 5.

An increase in the proportion of situational engagement meant an increase in the likelihood of getting higher grades in a math course (Table VII), which is in line with Hendolin's results about more frequent reporting of situational engagement showing up in grades [45]. Still, when the effect of the emotion residuals on the individual level on course success were tested, the effects of the situational engagement and learning situations were controlled inherently [58]. To elaborate, the individuals with the highest enhancers compared to all participants common mean, excluding the situational variation caused by situational engagement and learning situation, had higher chances to pass more courses and get higher grades on certain courses. This indicates that both the amount of situational engagement and the level of emotions regardless of situational engagement might be important factors for course success.

The third research question was about individual level emotions, gender, proportion of situational engagement, proportion of learning situations, passing certain courses, and their grades predicting student retention. The data supported only that higher individual level learning enhancing emotions and success in a chosen physics course would mean a more likely retention (Table VIII). Remembering, that the two chosen physics courses correlated strongly in both passing the course and in grades, this result is generalized to concern both chosen physics courses. Passing courses was not connected to retention like it earlier has [13], but in our study we had significantly limited the amount of courses to be observed. However, a novel and notable result is that the enhancing emotions not just had a significant positive effect on student retention, but that the OR concerning enhancers was much higher than the OR of the course grade, suggesting that the enhancing emotions could be a more powerful predictor of retention than the course grades.

A missing piece in the puzzle is does seeking higher situational enhancing emotions in learning situations lead to an increase in individual level enhancing emotions? This question was beyond our research design, and would be an important additive in the future. If it is accepted that a change in situational learning enhancing emotions would lead to a change in individual level learning enhancing emotions, situational engagement is definitely something to pursue. In practice, engaging learning situations have been sought and shown achieved, e.g., by replacing listening to lectures with culturally sensible and more active instructional methods [40,41], and making sure the chosen methods require deep learning strategies [42]. If one doubts that higher situational enhancing emotions lead to higher individual level enhancing emotions, the individual level emotions are still an important subject for some other individual level affective means, such as guidance and counseling, and overall interaction with the students.

VI. CONCLUSIONS

In this article, we explored the connections between learning situations, situational engagement and other emotions [22], gender, course success as grades and passed courses, and retention in HE physics education program. Learning situations and situational engagement were shown to connect to situational emotions independently in various ways (Tables II–IV), and individual level enhancing emotions in turn affected course success (Table V and VI) and retention in HE physics (Table VIII). The effect of learning situations on emotions was nested within gender (Table II–IV). We propose as novel results, that learning enhancing emotions could be a more important predictor of retention than grades (Table VIII).

We used a theoretical framework of optimal learning moment [22], which was rarely used in HE before. Since our results followed the line of previous research of using the model on a secondary level of education well, we suggest the OLM model could be used in HE research in the future too. We also showed one way to test the connection from situational engagement to science learning. The results suggest it exists especially through the learning enhancing emotions, which situational engagement boosts (Tables II, V, and VI).

In analyzing the hierarchical data, we used hierarchical linear analysis and penalized multinomial logistic regressions, which allowed us to follow an intriguing path through situational and individual levels of data from 20 Finnish first year HE physics students, and end up with situational as well as individual level outcomes. The analysis technique was compiled from several central sources and written out in detail to serve as an example for those aiming to use hierarchical analysis in, e.g., physics education research, even in smaller scale studies, and when the outcome variables are not on the lowest level of data hierarchy.

The main theoretical implication of this study is that the role of emotions in retention theories could be refined to be more prominent. We argued that emotions deserve a place among other dynamic variables, which was supported by our results. Emotions were shown to be not just an outcome, nor an independent predictor, but an interlocked, dynamic part of the student experience in and outside learning situations, leading to course success and retention (Fig. 3). This subject could benefit from more research having, e.g., larger samples, different cultural contexts, more detailed information about the situations, and data about change of emotions over time.

The practical implications of this study adhere to previous studies in that situational engagement and other emotions should be regarded seriously when designing learning situations, and that teaching or learning interactions, where emotions are regarded more instead of mere quantitative academic success, would be worth pursuing.

ACKNOWLEDGMENTS

The work of L. T. was funded by the Nano and molecular research unit and Faculty of Education, both from University of Oulu, Finland. L. T. wishes to thank her supervisor docent Saana-Maija Aho and colleague MSc Anu Tuomela for support during this research and writing process. Both are affiliated with the Nano and molecular research unit, University of Oulu, Finland.

APPENDIX A: ESM-questionnaire

The participants' situational variables were measured using experience sampling method, where a questionnaire consisting of the questions presented below, was sent to their personal cellphones several times a week at random times. Question number 2 was used to define, whether the participant was in a learning situation (response 2a) or in some other situation (responses 2b to 2e). Questions number 5 to 16 were used to measure the singular situational emotions, which were later summed as situational engagement (interest, challenge and skill), learning enhancers (enjoyment, successful, happy, confident, active), learning detractors (bored, confused), and learning accelerants (stress, anxiety).

1. Where are you? (*open answer*)
2. Were you
   (a) learning
   (b) at work
   (c) at a hobby
   (d) at leisure time
   (e) none of the above
3. Specify, what and how you were learning (*conditional: asked only if student responded 2(a)*)
   (a) I was at a physics lecture
   (b) I was at a supervised physics exercise
   (c) I was doing physics exercises without supervision
   (d) I was studying physics literature or other such material
   (e) I was doing physics lab work
   (f) I was learning physics in some other way than mentioned above
   (g) I was learning another subject than physics
   (h) none of the above
4. Are you
   (a) alone
   (b) together with someone
   (c) in a group (3–10 persons)
   (d) in a crowd (more than 10 persons)
5. Did you enjoy the thing you were doing? 1 = not at all … 5 = very much so
6. Was the thing you were doing challenging? 1 = not at all … 5 = very much so
7. Was the thing you were doing interesting? 1 = not at all … 5 = very much so
8. Do you feel skillful in what you were doing? 1 = not at all … 5 = very much so
9. Do you feel successful in what you were doing? 1 = not at all … 5 = very much so
10. How happy are you feeling? 1 = not at all … 5 = very much so
11. How confident are you feeling? 1 = not at all … 5 = very much so
12. How active are you feeling? 1 = not at all … 5 = very much so
13. How bored are you feeling? 1 = not at all … 5 = very much so
14. How confused are you feeling? 1 = not at all … 5 = very much so
15. How stressed are you feeling? 1 = not at all … 5 = very much so
16. How anxious are you feeling? 1 = not at all … 5 = very much so

APPENDIX B: HIERARCHICAL LINEAR MODELING RESULTS (ANALYSIS STEP 1)

In the first step of the analysis, three rounds of hierarchical linear modeling were executed. Each round had the same predictors (gender, learning situation,

TABLE IX. Hierarchical linear modeling results with enhancers as the outcome. All statistically significant parameter estimates (PE) are written in italic font. The −2LL was used to compare the same models with and without slope variance; italic font indicates better fit. The bolded column M2 is the best fitting model considering all L1, L2, and interaction terms and residuals.

		M0								M1										**M2**					
		PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	**PE**	SE	PE	SE	PE	SE
Fixed																									
Intercept	γ_{00}	*3.53*	0.11	*3.53*	0.11	*3.53*	0.11	*3.53*	0.11	*3.53*	0.11	*3.50*	0.20	*3.52*	0.13	*3.6*	0.2	*3.53*	0.11	**3.53**	0.11	*3.53*	0.11	*3.53*	0.11
L1																									
$Learning_{\mathrm{cwc}}$				*−0.44*	0.06	*−0.44*	0.08	*−0.46*	0.06	*−0.46*	0.08														
$Engagement_{\mathrm{cwc}}$								*0.35*	0.09	*0.33*	0.09									**0.33**	0.09	*0.35*	0.09	*0.33*	0.09
$Learning_{\mathrm{cgm}}$												*−0.46*	0.08	*−0.46*	0.08	*−0.5*	0.1								
$Engagement_{\mathrm{cgm}}$												*0.35*	0.09	*0.35*	0.09	*0.3*	0.1								
L2																									
Gender												0.04	0.23												
$LearningMean_{D40}$														0.03	0.20										
$EngagementMean_{D10}$																−0.1	0.2								
Interaction																									
$Learning_{\mathrm{cwc}}(Gender)$																		*−0.41* *−0.62*	0.09 (m) 0.15 (f)	**−0.41** **−0.63**	0.08 (m) 0.15 (f)	*−0.41* *−0.63*	0.06 (m) 0.12 (f)	*−0.41* *−0.63*	0.08 0.15
$Engagement_{\mathrm{cwc}}(Gender)$																		*0.34* 0.29	0.10 (m) 0.20 (f)						
$Learning_{\mathrm{cwc}} * Engagement_{\mathrm{cwc}}$																								0.05	0.20
Random																									
L1 residual	$\mathrm{var}(e_{ij})$	*0.34*	0.02	*0.30*	0.02	*0.28*	0.02	*0.29*	0.02	*0.28*	0.02	*0.28*	0.02	*0.28*	0.02	*0.3*	0.0	*0.28*	0.02	**0.28**	0.02	*0.29*	0.02	*0.28*	0.02
L2 residual	$\mathrm{var}(u_{0j})$	*0.21*	0.07	*0.21*	0.07	*0.21*	0.07	*0.21*	0.07	*0.21*	0.07	*0.18*	0.07	*0.18*	0.07	*0.2*	0.1	*0.21*	0.07	**0.21**	0.07	*0.21*	0.07	*0.21*	0.07
Slope variance	$\mathrm{var}(u_{1j})$					*0.06*	0.04			*0.05*	0.03	0.05	0.04	0.05	0.04	0.1	0.0	0.05	0.03	**0.05**	0.03			0.05	0.03
ICC		0.38																							
−2LL		827.26		774.75		*767.99*		762.01		*756.98*		754.69		754.95		754.5		758.15		*757.03*		761.35		758.40	

TABLE X. Hierarchical linear modeling results with detractors as the outcome. All statistically significant parameter estimates (PE) are written in italic font. The −2LL was used to compare same models with and without slope variance; italic font indicates better fit. The bolded column M2 is the best fitting model considering all L1, L2, and interaction terms and residuals.

		M0								M1										**M2**					
		PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	PE	SE	**PE**	SE	PE	SE	PE	SE
Fixed																									
Intercept	γ_{00}	*2.14*	0.11	*2.14*	0.11	*2.14*	0.11	*2.14*	0.11	*2.14*	0.11	*1.98*	0.20	*2.16*	0.13	*2.30*	0.16	*2.14*	0.11	*2.14*	0.11	*2.14*	0.11	*2.14*	0.11
L1																									
$Learning_{\mathrm{cwc}}$				*0.56*	0.07	*0.56*	0.10	*0.58*	0.07	*0.58*	0.09														
$Engagement_{\mathrm{cwc}}$								*−0.32*	0.11	*−0.30*	0.11									**−0.30**	0.11	*−0.32*	0.11	*−0.29*	0.11
$Learning_{\mathrm{cgm}}$												*0.59*	0.09	*0.59*	0.09	*0.59*	0.10								
$Engagement_{\mathrm{cgm}}$												*−0.31*	0.11	*−0.30*	0.11	*−0.32*	0.11								
L2																									
$Gender$												0.21	0.24												
$LearningMean_{D40}$														-0.06	0.21										
$EngagementMean_{D10}$																-0.28	0.20								
Interaction																									
$Learning_{\mathrm{cwc}}(Gender)$																		*0.56* *0.67*	0.11 (m) 0.20 (f)	*0.57* *0.63*	0.11 (m) 0.20 (f)	*0.57* *0.65*	0.08 (m) 0.15 (f)	*0.57* *0.64*	0.11 (m) 0.20 (f)
$Engagement_{\mathrm{cwc}}(Gender)$																		*−0.22* -0.63	0.12 (m) 0.25 (f)						
$Learning_{\mathrm{cwc}} * Engagement_{\mathrm{cwc}}$																								-0.34	0.24
Random																									
L1 residual	$\mathrm{var}(e_{ij})$	*0.52*	0.04	*0.45*	0.03	*0.43*	0.03	*0.44*	0.03	*0.42*	0.03	*0.42*	0.03	*0.42*	0.03	*0.42*	0.03	*0.42*	0.03	*0.42*	0.03	*0.442*	0.031		
L2 residual	$\mathrm{var}(u_{0j})$	*0.20*	0.07	*0.20*	0.07	*0.20*	0.07	*0.20*	0.07	*0.20*	0.07	*0.19*	0.07	*0.19*	0.07	*0.17*	0.06	*0.20*	0.07	*0.20*	0.07	*0.204*	0.072		
Slope variance	$\mathrm{var}(u_{1j})$					*0.09*	0.06			*0.08*	0.06	0.08	0.06	0.08	0.06	0.08	0.06	0.09	0.06	*0.09*	0.06				
ICC		0.28																							
−2LL		1005.86		947.46		*942.03*		941.28		*936.82*		935.27		936.20		934.55		936.45		*937.93*		942.81		936.96	

TABLE XI. First half of hierarchical linear modeling results with LOG10(accelerants) as the outcome. All statistically significant parameter estimates (PE and 10^{PE}) are in italic font. The −2LL was used to compare same models with and without slope variance; italic font indicates better fit.

		M0									M1					
		PE	10^{PE}	SE	PE	10^{PE}	SE	PE	10^{PE}	SE	PE	10^{PE}	SE	PE	10^{PE}	SE
Fixed																
Intercept	γ_{00}	*0.31*	*2.06*	0.03	*0.31*	*2.06*	0.03	*0.31*	*2.06*	0.03	*0.31*	*2.06*	0.03	*0.28*	*1.89*	006
L1																
$Learning_{\mathrm{cwc}}$					*0.11*	*1.28*	0.02	*0.11*	*1.28*	0.02	*0.11*	*1.28*	0.02			
$Engagement_{\mathrm{cwc}}$											0.00	1.01	0.02			
$Learning_{\mathrm{cgm}}$														*0.11*	*1.29*	002
$Engagement_{\mathrm{cgm}}$																
L2																
Gender														0.05	*1.11*	007
Random																
Residual	$\mathrm{var}(e_{ij})$	*0.03*	*1.06*	0.00	*0.02*	*1.05*	0.00	*0.02*	*1.05*	0.00	*0.02*	1.05	0.00	*0.02*	*1.05*	000
	$\mathrm{var}(u_{0j})$	*0.02*	*1.05*	0.01	*0.02*	*1.05*	0.01	*0.02*	*1.05*	0.01	*0.02*	1.05	0.01	*0.02*	*1.05*	0.01
Slope variance	$\mathrm{var}(u_{1j})$							*0.00*	*1.00*	0.00						
ICC		0.44														
−2LL		−302.80			−341.62			*−342.78*			−336.05			−340.00		

TABLE XII. Second half of hierarchical linear modeling results with LOG10(accelerants) as the outcome. All statistically significant parameter estimates (PE and 10^{PE}) are written with cursive. The −2LL was used to compare same models with and without slope variance; italic font indicates better fit. The bolded column M2 is the best fitting model considering all L1, L2, and interaction terms and residuals.

								M2								
		PE	10^{PE}	SE	PE	10^{PE}	SE	**PE**	10^{PE}	SE	PE	10^{PE}	SE	PE	10^{PE}	SE
Fixed																
Intercept	γ_{00}	*0.34*	*2.17*	0.04	*0.38*	*2.40*	0.05	*0.31*	*2.06*	0.03	*0.31*	*2.06*	0.03	*0.31*	*2.05*	0.03
L1																
$Learning_{\mathrm{cgm}}$		*0.11*	*1.28*	0.02	*0.11*	*1.29*	0.02									
L2																
$LearningMean_{D40}$		−0.06	0.86	0.06												
$EngagementMean_{D10}$					−0.11	0.77	0.06									
Interaction																
$Learning_{\mathrm{cwc}}(Gender)$								**0.10**	*1.27*	0.02 (m)	*0.10*	*1.27*	0.02 (m)	*0.10*	*1.27*	0.02 (m)
								0.13	*1.34*	0.03 (f)	*0.12*	*1.32*	0.04 (f)	*0.12*	*1.33*	0.03 (f)
$Learning_{\mathrm{cwc}} * Engagement_{\mathrm{cwc}}$														0.07	*1.19*	0.05
Random																
L1 residual	$\mathrm{var}(e_{ij})$	*0.02*	1.05	*0.00*	*0.02*	*1.05*	0.00	*0.02*	*1.05*	0.00	*0.02*	*1.05*	0.00	*0.02*	*1.05*	0.00
L2 residual	$\mathrm{var}(u_{0j})$	*0.02*	1.04	*0.01*	*0.02*	*1.04*	0.01	*0.02*	*1.05*	0.01	*0.02*	*1.05*	0.01	*0.02*	*1.05*	0.01
Slope variance	$\mathrm{var}(u_{1j})$										0.00	*1.01*	0.00			
−2LL		−340.37			−342.66			**−337.28**			*−338.57*			−335.18		

situational engagement, the mean aggregations and interaction terms of the latter two, and their nesting within gender). Results of this phase of modeling for the three different outcomes are presented in their entirety in the following tables: for learning enhancers in Table IX, for learning detractors in Table X and for 10 based logarithmic values of learning accelerants in Tables XI–XII.

[1] The Organization for Economic Cooperation and Development, *Education at a Glance 2019* (OECD, Paris, France, 2019), p. 520.

[2] The Organization for Economic Cooperation and Development, *Distribution of graduates and entrants by Field* (OECD, Paris, France, 2018).

[3] The Organization for Economic Cooperation and Development, *Education at a Glance 2020* (OECD, Paris, France, 2020), p. 476.

[4] E. R. Kahu, Framing student engagement in higher education, Studies Higher Educ. **38**, 758 (2013).

[5] E. Kahu, C. Stephens, L. Leach, and N. Zepke, Linking academic emotions and student engagement: mature-aged distance students' transition to university, J. Further Higher Educ. **39**, 481 (2015).

[6] C. Mega, L. Ronconi, and R. De Beni, What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic Achievement, J. Educ. Psychol. **106**, 121 (2014).

[7] A. Bager-Elsborg, K. J. Herrmann, R. Troelsen, and L. Ulriksen, Are leavers different from stayers?, Uniped **42**, 139 (2019).

[8] J. P. Bean, Student attrition, intentions, and confidence: Interaction effects in a path model. Part I. The twenty-three variable model, in *Proceedings of the Annual Meeting of the American Educational Research Association, Los Angeles, 1981* (ERIC Document Reproduction Service # HE 013 983) (1981).

[9] V. Tinto, Dropout from higher education: A theoretical synthesis of recent research, Rev. Educ. Res. **45**, 89 (1975).

[10] J. Bean and S. B. Eaton, The psychology underlying successful retention practices, J. College Student Retention: Res. Theory Pract. **3**, 73 (2001).

[11] J. P. Bean, Dropouts and turnover: The synthesis and test of a causal model of student attrition, Res. High. Educ. **12**, 155 (1980).

[12] E. T. Pascarella, Student-faculty informal contact and college outcomes, Rev. Educ. Res. **50**, 545 (1980).

[13] D. Tafreschi and P. Thiemann, Doing it twice, getting it right? The effects of grade retention and course repetition in higher education, Econ. Educ. Rev. **55**, 198 (2016).

[14] A. González, M.-V. C. M. V. C. Fernández, and P.-V. V. Paoloni, Hope and anxiety in physics class: Exploring their motivational antecedents and influence on metacognition and performance, J. Res. Sci. Teach. **54**, 558 (2017).

[15] F. T. Villavicencio and A. B. Bernardo, Positive academic emotions moderate the relationship between self-regulation and academic achievement, Br. J. Educ. Psychol. **83**, 329 (2013).

[16] S. Cwik and C. Singh, Students' sense of belonging in introductory physics course for bioscience majors predicts their grade, Phys. Rev. Phys. Educ. Res. **18**, 010139 (2022).

[17] L. J. Sax, K. J. Lehman, R. S. Barthelemy, and G. Lim, Women in physics: A comparison to science, technology, engineering, and math education over four decades, Phys. Rev. Phys. Educ. Res. **12**, 020108 (2016).

[18] V. Tinto, *Leaving college: Rethinking the causes and cures of student attrition*, 2nd ed. (University of Chicago Press, Chicago, 1994).

[19] M. K. Udo, G. P. Ramsey, and J. V. Mallow, Science anxiety and gender in students taking general education science courses, J. Sci. Educ. Technol. **13**, 435 (2004).

[20] M. K. Udo, G. P. Ramsey, S. Reynolds-Alpert, and J. V. Mallow, Does physics teaching affect gender-based science anxiety?, J. Sci. Educ. Technol. **10**, 237 (2001).

[21] R. S. Barthelemy and A. V. Knaub, Gendered motivations and aspirations of university physics students in Finland, Phys. Rev. Phys. Educ. Res. **16**, 010133 (2020).

[22] B. Schneider, J. Krajcik, J. Lavonen, K. Salmela-Aro, M. Broda, J. Spicer, J. Bruner, J. Moeller, J. Linnansaari, K. Juuti, and J. Viljaranta, Investigating optimal learning moments in U.S. and Finnish science classes, J. Res. Sci. Teach. **53**, 400 (2016).

[23] F. Fujita, E. Diener, and E. Sandvik, Gender differences in negative affect and well-being: The case for emotional intensity, J. Personality Soc. Psychol. **61**, 427 (1991).

[24] K. M. Whitcomb and C. Singh, For physics majors, gender differences in introductory physics do not inform future physics performance, Eur. J. Phys. **41**, 065701 (2020).

[25] J. Rosenberg, P. N. Beymer, V. Phun, and J. Schmidt, How does situational engagement vary between learners, situations, and classrooms? Findings from the use of intensive longitudinal methods and cross-classified, multi-level models, 10.31219/osf.io/pj2v8 (2020).

[26] K. Salmela-Aro, K. Upadyaya, P. Cumsille, J. Lavonen, B. Avalos, and J. Eccles, Momentary task-values and expectations predict engagement in science among Finnish and Chilean secondary school students, Int. J. Psychol. **56**, 415 (2021).

[27] B. Van Dusen and J. Nissen, Modernizing use of regression models in physics education research: A review of hierarchical linear modeling, Phys. Rev. Phys. Educ. Res. **15**, 020108 (2019).

[28] D. Robotham and C. Julian, Stress and the higher education student: A critical review of the literature, J. Further Higher Educ. **30**, 107 (2006).

[29] S. Zirkel, J. A. Garcia, and M. C. Murphy, Experience-sampling research methods and their potential for education research, Educ. Res. **44**, 7 (2015).

[30] A. W. Astin, Student involvement: A developmental theory for higher education, J. Coll. Student Dev. **40**, 518 (1999).

[31] J. P. Bean, The synthesis of a theoretical model of student attrition, in *Proceedings of the Annual Meeting of the American Educational Research Association, Los Angeles, CA, 1981* (ERIC Document Reproduction Service No ED 202444), https://eric.ed.gov/?id=ED202444.

[32] A. F. Cabrera, A. Nora, and M. B. Castañeda, College Persistence: Structural equations modeling test of an integrated model of student retention, J. Higher Educ. **64**, 123 (1993).

[33] A. F. Cabrera, M. B. Castañeda, A. Nora, and D. Hengstler, The convergence between two theories of college persistence, J. Higher Educ. **63**, 143 (1992).

[34] E. T. Pascarella, How college affects students: Ten directions for future research, J. Coll. Student Dev. **47**, 508 (2006).

[35] M. Tight, Student retention and engagement in higher education, J. Further Higher Educ. **44**, 689 (2020).

[36] O. Aljohani, A Comprehensive review of the major studies and theoretical models of student retention in higher education, Higher Educ. Studies **6**, 1 (2016).
[37] B. O. Barefoot, Higher education's revolving door: Confronting the problem of student drop out in US colleges and universities, Open Learning: The Journal of Open, Distance and e-Learning **19**, 9 (2004).
[38] F. Manyanga, A. Sithole, and S. M. Hanson, Comparison of student retention models in undergraduate education from the past eight decades, J. App. Learning Higher Educ. **7**, 30 (2017), https://www.missouriwestern.edu/appliedlearning/wp-content/uploads/sites/116/2018/11/JALHE-7.pdf.
[39] D. R. Johnson, T. H. Wasserman, N. Yildirim, and B. A. Yonai, Examining the effects of stress and campus climate on the persistence of students of color and white students: An application of Bean and Eaton's psychological model of retention, Res. High. Educ. **55**, 75 (2014).
[40] J. Inkinen, K. Christopher, S. Barbara, J. Kalle, K. Joseph, L. Jari, and S.-A. Katariina, Science classroom activities and student situational engagement, Int. J. Sci. Educ. **41**, 316 (2019).
[41] J. Inkinen, C. Klager, K. Juuti, B. Schneider, K. Salmela-Aro, J. Krajcik, and J. Lavonen, High school students' situational engagement associated with scientific practices in designed science learning situations, Sci. Educ. **104**, 667 (2020).
[42] D. Yang, Engaging Secondary School Students in Science Learning through a Massive Open Online Course (MOOC), Ph.D. thesis, the University of Helsinki, 2021.
[43] S. Lehtamo, K. Juuti, J. Inkinen, and J. Lavonen, Connection between academic emotions *in situ* and retention in the physics track: Applying experience sampling method, Int. J. STEM Educ. **5**, 25 (2018).
[44] X. Tang, J. Lavonen, B. Schneider, J. Krajcik, and K. Salmela-Aro, High school students' optimal learning moments: A network analysis approach, in *Proceedings of the 2020 Annual Meeting of the American Educational Research Association*. Retrieved 6/1/2021, from the AERA Online Paper Repository.
[45] I. Hendolin, Exploring optimal learning moments at tutorial sessions, in *Proceedings of PER Conf. 2016, Sacramento, CA*, edited by D. L. Jones, L. Ding, and A. Traxler (American Association of Physics Teachers, Sacramento, CA, 2016), pp. 144–147, 10.1119/perc.2016.pr.031.
[46] K. A. Douglas, M. S. Yale, D. E. Bennett, M. P. Haugan, and L. A. Bryan, Evaluation of colorado learning attitudes about science survey, Phys. Rev. ST Phys. Educ. Res. **10**, 020128 (2014).
[47] Vipunen—Finnish National Agency for Education's Statistic Services, Universities' new students, students and degrees.
[48] S. L. Britt, D. A. Ammerman, S. F. Barrett, and S. Jones, Student loans, financial stress, and college student retention, J. Student Financial Aid, **47**, 25 (2017).
[49] The Social Insurance Institution of Finland, Study Progress in Higher Education. Retrieved July 9, 2021, https://www.kela.fi/web/en/study-progress-higher-education.
[50] Universities Act 24.7.2009/558, https://www.finlex.fi/en/laki/kaannokset/2009/20090558.
[51] T. L. Grande, Fisher's Exact Test in SPSS (2015), https://youtu.be/6MbTlee-He8.
[52] BrunelASK, Reliability test: Compute Cronbach's alpha using SPSS (2015), https://youtu.be/OdIhhtg-3LE.
[53] BrunelASK, Reliability test: Interpret Cronbach's alpha output in SPSS (2015), https://youtu.be/pVfByfoQ1lU.
[54] J. Pallant, *SPSS Survival Manual—A Step by Step Guide to Data Analysis Using SPSS*, 4th ed. (Allen and Unwin, New South Wales, Australia, 2011).
[55] J. J. Hox, M. Moerbeek, and R. van de Schoot, *Multilevel Analysis: Techniques and Applications*, 3rd ed., Quantitative methodology series (Routledge, New York, NY, 2018).
[56] G. Gilles, Little's Missing Completely at Random (MCAR) Test—SPSS (2011), https://youtu.be/22aR9ruSig4.
[57] S. van Buuren, Multiple imputation of multilevel data, *Handbook of Advanced Multilevel Analysis*, edited by J. J. Hox and J. K. Roberts (Routledge, New York, 2010), pp. 173–196, 10.4324/9780203848852.ch10.
[58] M. A. Griffin, Interaction between and situations: Using HLM procedures to estimate reciprocal relationships, J. Management **23**, 759 (1997).
[59] T. A. B. Snijders and R. J. Bosker, *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 2nd ed. (Sage Publications, Thousand Oaks, CA, 2012).
[60] D. Becker, W. Breustedt, and C. I. Zuber, Surpassing simple aggregation: Advanced strategies for analyzing contextual-level outcomes in multilevel models, Methods, data, analyses **12**, 31 (2018).
[61] C. K. Enders and T. Tofighi, Centering predictor variables in cross-sectional multilevel models: A new look at an old issue, Psychological methods **12**, 121 (2007).
[62] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2017).
[63] G. Heinze, M. Ploner, and L. Jiricka, logistf: Firth's Bias-Reduced Logistic Regression (2020), r package version 1.24, https://rdocumentation.org/packages/logistf/versions/1.24.
[64] K. Ioannis, brglm2: Bias Reduction in Generalized Linear Models (2021), r package version 0.7.1.
[65] D. McNeish and L. M. Stapleton, Modeling clustered data with very few clusters, Multivariate Behav. Res. **51**, 495 (2016).
[66] B. A. Bell, G. B. Morgan, J. A. Schoeneberger, J. D. Kromrey, and J. M. Ferron, How low can you go?, Methodology **10**, 1 (2014).
[67] X. Wang, Firth logistic regression for rare variant association tests, Front. Genetics **5**, 187 (2014).
[68] D. Firth, Bias reduction of maximum likelihood estimates, Biometrika **80**, 27 (1993).
[69] S. B. Bull, C. Mak, and C. M. T. Greenwood, A modified score function estimator for multinomial logistic regression in small samples, Computational Statistics and Data Analysis **39**, 57 (2002).
[70] J. Lorah, Effect size measures for multilevel models: definition, interpretation, and TIMSS example, Large-scale Assess. Educ. **6**, 8 (2018).

[71] K. P. Burnham and D. R. Anderson, Multimodel inference: Understanding AIC and BIC in model selection, Sociological Methods Res. **33**, 261 (2004).
[72] J. Cohen, A Power Primer, Psychol. Bull. **112**, 155 (1992).
[73] L. Ulriksen, L. M. Madsen, and H. T. Holmegaard, What do we know about explanations for drop out/opt out among young people from STM higher education programmes?, Studies Sci. Educ. **46**, 209 (2010).
[74] Y. Li and C. Singh, Do female and male students' physics motivational beliefs change in a two-semester introductory physics course sequence?, Phys. Rev. Phys. Educ. Res. **18**, 010142 (2022).
[75] R. B. King and M. J. M. Gaerlan, High self-control predicts more positive emotions, better engagement, and higher achievement in school, Eur. J. Psychol. Educ. **29**, 81 (2014).
[76] J. R. Rudland, C. Golding, and T. J. Wilkinson, The stress paradox: How stress can be good for learning, Med. Educ. **54**, 40 (2020).