Self-efficacy and conceptual knowledge in quantum mechanics during teaching reforms and the COVID-19 pandemic

Elina Palmgren,[1,*] Kimmo Tuominen,[1,†] and Inkeri Kontro[2,‡]
[1]Department of Physics, University of Helsinki, POB 64, FI-00014 Helsinki, Finland
[2]Physics Unit, Tampere University, FI-33720 Tampere, Finland

(Received 22 December 2021; accepted 10 August 2022; published 4 October 2022)

Physics instruction is often unable to support students' self-efficacy. The remote teaching brought on by the COVID-19 pandemic has also affected learning. We surveyed an introductory quantum mechanics course for three years during a transition into the spin first approach, adapting the student-centered *prime-time learning* model and using it through the remote teaching during the pandemic. Prime-time learning includes weekly meetings where students and instructors discuss in a small group, and the assessment is based on exercises, group work, and self-assessment. We show that this teaching method improved students' self-efficacy. Students' conceptual knowledge post teaching remained high throughout the teaching reform, as measured by an abbreviated Quantum Mechanics Concept Assessment test. We also find that the prime-time model is remarkably stable during remote teaching: in contrast to many other studies, we did not see a decline in conceptual learning outcomes or self-efficacy in remote teaching during the COVID-19 pandemic.

DOI: 10.1103/PhysRevPhysEducRes.18.020122

I. INTRODUCTION

Quantum mechanics (QM) is often regarded as a notoriously difficult topic to learn. For the instructors of QM this reputation is an unfortunate one, as it has been shown in previous literature that these kinds of presumptions are likely to reduce students *self-efficacy* and thus affect their learning [1]. Self-efficacy refers to the beliefs a person has about their ability to succeed in a given topic [2]. These beliefs are highly situational, and they uniquely predict study success [3,4]. Self-efficacy can be supported by experiences of mastery, social persuasion, and by reducing stress and negative feelings [2], which also are key components to successful learning. This is particularly important in QM, because studies show that on QM courses, students are often valued based on their ability to calculate, which leads to a culture that can hurt students, particularly those who are not perceived as typical physics majors [5].

The aforementioned culture that emphasizes calculation skills is strongly present in the instructional approach QM courses have traditionally used, namely, the "position first" approach. This approach often starts from the introduction of wave functions, and students begin with calculations in continuous bases and infinite-dimensional systems [6].

During recent years, a new approach to teaching QM has emerged, aiming to reduce the emphasis on calculations. This novel approach, the "spin first" approach, focuses on using spin-$1/2$ particles and sequential Stern-Gerlach experiments [7] as a context to introduce the postulates of QM [6,8,9]. Even if the research in the effectiveness of the spin first approach is limited, the results so far have been promising [6,9].

In order to support student learning as much as possible, the spin first instructional approach should be applied together with active learning teaching strategies [10,11]. One such teaching strategy, *prime-time learning* [12], has been pioneered by University of Jyväskylä, Finland, and adapted for use in many other Finnish universities [13]. The method improves student commitment and supports learning by offering student groups individualized support from teachers [12]. Prime-time learning has also proven resilient during the challenges brought by the emergency remote teaching in the COVID-19 pandemic [13].

At the University of Helsinki, we have taught an introductory course in QM combining the spin first approach and the prime-time method. The teaching reform preceded the shift to remote teaching due to the COVID-19 pandemic, which enabled us to study whether students' performance changed due to remote teaching.

In this paper, we discuss how students' self-efficacy can be supported by prime-time learning also during remote teaching, and how the spin first approach is suited for an introductory university course. To this end, we, on the one

*elina.palmgren@helsinki.fi
†kimmo.i.tuominen@helsinki.fi
‡inkeri.kontro@tuni.fi

hand, studied students' self-efficacy and, on the other hand, their conceptual understanding and knowledge profiles. Our research questions were

1. Does prime-time learning affect the self-efficacy beliefs in quantum mechanics, and what was the effect of remote teaching?
2. Does prime-time learning affect the learning of QM concepts, and what was the effect of remote teaching?
3. What kind of QM knowledge profiles emerge from the data?
4. How are knowledge profiles and self-efficacy related?

To address these research questions, we begin by describing the background in Sec. II, where we introduce the concept of self-efficacy, the spin first approach, and the Finnish university education system as well as the prime-time method. In Sec. III, we describe the study methods, and in Sec. IV, we present the results of the study and discuss the implications by research question. We conclude with a summary in Sec. V.

II. BACKGROUND

A. Self-efficacy

Student beliefs regarding their ability to perform are collectively called self-efficacy. These beliefs are built up through experiences of successes, through interpersonal contacts (social models and persuasion) and by reducing stress and negative feelings [2,14]. For example, mastery experiences predict self-efficacy in science [15]. On the other hand, self-efficacy beliefs predict success in studies [3,4], and they are precursors of student interest and choice of major subject [16]. For physics students, self-efficacy is highly aligned with physics identity [17], and physics identity predicts a choice of physics as a career [18]. Self-efficacy can be supported with teaching. For example, using self-assessment for administering grades links to higher self-efficacy for mathematics students [19]. However, university physics instruction has often negative effect on self-efficacy beliefs [20,21].

In addition to physics identity, gender is an important factor in self-efficacy. Male physics students generally have higher self-efficacy than female students with similar grades [22], and recognition from teachers influences women's and men's self-efficacy in different ways [17]. Women studying physics in Finland report more anxiety and lower self-efficacy than men [23]. However, studies often view physics as a monolith and only sparse literature exists on the relationship between the subfields of physics, although many physicists feel that subfields of physics are valued differently [24]. Physicists from certain specializations such as applied physics may not even feel like proper physicists. This division is experienced by both genders [24]. To our knowledge, no studies have probed the connection between self-efficacy beliefs and physics subfield.

In addition, research shows that particularly on quantum mechanics courses, physics students are often valued based on their ability to calculate, which leads to a culture that can hurt students who are not perceived as typical physics majors [5]. However, changing the culture of the courses can be difficult, as students may not readily accept teaching styles which are different from their expectations [25].

B. Spin first

Quantum mechanics (QM) is a highly interesting and motivating, but abstract and difficult topic for university physics students. Often QM courses start from the Schrödinger equation and wave functions, and introduce problems related to the energy and position of particles in potential wells. This instructional approach is called the position first approach.

Recently, a new approach, called the spin first approach, has emerged in teaching QM. This approach utilizes spin-$1/2$ particles and sequential Stern-Gerlach experiments [7] as a context to introduce the postulates of quantum mechanics within a mathematically simple two-state system [6,8,9].

The idea of the spin first approach is to focus on introducing quantum mechanical concepts without tedious calculations involving differential operators in the infinite dimensional Hilbert space. Rather, the concepts are discussed in terms of discrete systems, where states are represented using the Dirac notation and observables are represented by operators with simple action on the states. These are explicitly connected to their vector and matrix representations. Examining the simplest systems possible, the spin-$1/2$ particles, we end up using two-component state vectors and 2×2 matrices. This is to be contrasted with the traditional position first approach, where students need to do calculations in continuous bases and infinite-dimensional systems from the very beginning.

Even if the position and momentum representations of the position first approach are often regarded more intuitive for students, the previous research shows that using the spin first approach may improve their understanding of certain topics, such as quantum measurement and probability [6,9]. In addition, using the state representation supports student reasoning [26,27].

Furthermore, in the Finnish context, the spin first approach has been applied also in extracurricular high school studies [28], and qualitative treatment of spin systems was recently introduced to the Finnish high school physics curriculum [29]. Understanding these systems becomes thus vitally important for new student groups such as preservice teachers.

C. Finnish university education and prime-time learning

The Finnish education system is renowned for producing high learning outcomes in primary school especially in mathematics, science, and reading comprehension [30,31].

TABLE I. Summary of the teaching reforms by year.

Year	Approach	Teaching modality	Textbook	Assessment	Data
2018	Position first	Active onsite lectures	[37]	Exercises and exam	SE
2019	Spin first	Active onsite lectures	[38]	Exercises and exam	SE and QMCA and major
2020	Spin first	Onsite prime time	[38]	Exercises, group work and self-assessment	SE and QMCA and major
2021	Spin first	Remote prime time	[38]	Exercises, group work and self-assessment	SE and QMCA and major
2022	Spin first	Remote prime time	[38]	Exercises, group work and self-assessment	SE and QMCA and major

In high school, students study a broad collection of subjects, but focus on a few on which they take matriculation exams. University students choose their major subject when applying, and in science, minor subjects are in related fields. Engineering students generally attend technical universities. As a consequence, in calculus-based physics courses in research universities, virtually all students are science majors and the majority are physics majors.

Physical science students choose their study track early in their studies, usually in the first or second year. The study tracks at the University of Helsinki are (experimental) physics, theoretical physics, physics with broad science orientation, meteorology, and astronomy. Preservice physics teachers study mostly the same courses as students on the experimental physics study track. The mode of teaching is active and, recently, several courses have adapted the prime-time approach as an alternative to lectures and individual work.

The prime-time model uses four steps of learning: *Principles*, or the individual study time where students watch lecture videos and read course material; *practice*, with collaborative problem solving; *problems*, meaning traditional exercises which are solved either individually or in groups; and prime time, where the small group that has practiced together meets with the teacher and discusses their questions and problems. Previous research shows that the accountability students feel both towards their group members and the teacher reduces drop-out, and learning outcomes are at least as good as with other active learning methods [12]. While research on the prime-time learning method is still limited in scope, versions of it have been adopted in many physics courses in several Finnish universities [13].

The value in prime-time learning is the personal connection between the instructor and the student groups. The importance of social connections for learning has long been known. For example, good integration into the social network correlates with learning [32,33]. The COVID-19 pandemic has made this even more visible: the lack of contacts has been detrimental in particular to students in the beginning of their physics studies [34], and remote teaching has been detrimental to the self-efficacy of physics students in general [35].

D. Teaching reforms

The course examined in this study, called Basics of Quantum Physics, is an introductory course in theoretical physics at the University of Helsinki. The course is compulsory for all students of physics, theoretical physics, and meteorology, as well as for preservice physics teachers.

During recent years, many teaching reforms have taken place on the course. The course was taught in the position first approach until 2018, and in spin first from 2019 onward. In 2018–2019, the course had active lectures, and from 2020, prime-time learning was implemented. The formative assessment of 2020–2022 was augmented by students continuous self-assessment using the DISA platform [36]. Furthermore, due to the COVID-19 pandemic, the teaching was remote in 2021–2022. A summary of the changes is presented in Table I. In all years, there were also tutored exercise sessions to support learning.

III. METHODS

A. Questionnaires

To study the self-efficacy beliefs specifically in the context of quantum mechanics, we devised a short questionnaire loosely based on the questions by Bailey *et al.* [39]. The statements were formulated for a quantum mechanics course, and originally contained statements on the ability to understand the lectures, reading material, the mathematics, and the ability to perform in the exercises and exam to the student's own satisfaction. The reason to formulate the statements to include students' own satisfaction was to account for the different motivation levels of students and to support a mindset in which grades are not the most important outcome of the course. During the study, the statement referring to the exam was dropped, as from 2020, there was no exam. The statements are presented in Table II, and the validation of the questionnaire is presented in Sec. III C 1. The self-efficacy questions were answered on a 5-point Likert scale (strongly disagree, 1, to strongly agree, 5).

From 2019 and the transition to the spin first approach, we also surveyed conceptual knowledge post teaching. To do this, we used an abbreviated translation of QMCA (version 6.2.2) [40–42]. The original QMCA questionnaire includes 38 multiple-choice questions that are designed to measure students' understanding of quantum mechanical concepts in both the spatial wave function and the spins contexts. From this set of questions, we selected 13 questions, the topics of which were covered on Basics of Quantum Physics. These questions include seven problems in the spins context, three in the wave function context,

TABLE II. Self-efficacy statements, adapted from Ref. [39].

No.	Statement
SE1	I understand the concepts of quantum mechanics when I read.
SE2	I understand the concepts of quantum mechanics when I attend a lecture or prime-time meeting.
SE3	I understand the mathematics used in quantum mechanics.
SE4	My performance in exercises satisfied me.

and three problems about measurements of fictitious quantum mechanical observables called color and size. Six of these questions form three isomorphic pairs (questions 1 and 11, 2 and 12, and 3 and 13) that ask practically the same question but in the spins and wave function contexts. A summary of the contents of the post-test questions and their corresponding QMCA questions are presented in Table III. The post-test questions were translated into Finnish by the authors. After this, the Finnish questions were translated back into English by a native English speaker, who is fluent in Finnish. The minor differences in wording were resolved by discussion. The data were scored based on whether the answer was correct or not.

B. The sample and data collection

Each year at the end of the course, students were asked to fill up a course feedback form, which was a familiar procedure from previous courses. The self-efficacy questions were included in this feedback form. Students were granted exercise points corresponding to points from a single calculation problem. They were asked for informed consent of participation in the research, and exercise credit was given regardless of whether students gave consent to participate in research.

The dataset of 2018 comprises 50 students. Data collected from this course has also been used in Ref. [43].

In 2019–2022, as a part of the last course exercise sheet, students were also asked to take the abbreviated QMCA post-test. The post-test was administered using an electronic questionnaire on the Moodle platform of Basics of Quantum Physics, and exercise credit was given as described above. The test was not proctored, but students were asked not to use any course material in answering the questions and at the end of the questionnaire they were asked if they had used any additional material. Only answers from the students who reported that they had not used materials in answering the questions were used in analysis.

In addition, in 2019–2022 students were asked to report their major subject. This question was included either in a background questionnaire at the beginning of the course or in the feedback questionnaire at the end of the course.

In 2019–2022, only the answers from the students who gave their consent to using all the collected data were included in the dataset. From the students who had given the consent on their data being used in research we selected only those who had answered all the post-test questions. In addition, we had to omit one response from the analysis because the Rasch model introduced below was not able to treat it. This respondent had answered 5 to self-efficacy questions SE1–SE3 and left SE4 blank, resulting in a too extreme answering pattern for the model to handle. This brought the number of students in 2019–2022 to 222. The dataset includes two respondents who did not answer all self-efficacy questions (one each for SE1 and SE2).

All study participants were consenting adults. The research did not involve intervention in the physical

TABLE III. Post-test multiple-choice questions, their corresponding QMCA question numbers, and the conceptual content of each question. The descriptions of the question contents are adapted from Ref. [40].

No.	QMCA	Question summary
1	22	Spin-1/2 particle; What is the maximum value that can be measured from a superposition state?
2	23	Spin-1/2 particle; What is the normalized state after the most probable value is measured?
3	24	Spin-1 particle; What are the possible values of repeating measurements for noncommuting operators $\hat{S}_x$ and $\hat{S}_z$?
4	27	Spin-1/2 particle; What is the state after a measurement of $\hat{S}_z$?
5	28	Spin-1/2 particle; Does $\hat{Q}\|\psi\rangle$ describe the state after a measurement of $\hat{Q}$? (T/F question)
6	29	Spin-1/2 particle; Is there a definite spin value for a superposition state? (T/F question)
7	30	Spin-1/2 particle; Does the addition of a relative phase affect the probability of measuring spin?
8	34	ColorTron and SizeUp; Is a property preserved in immediate subsequent measurements?
9	35	ColorTron and SizeUp; Choose a valid representation of the state given measurement results on the state.
10	36	ColorTron and SizeUp; What are the possible values of repeating measurements for commuting operators?
11	1	Wave function; What is the maximum value that can be measured from a superposition state?
12	2	Wave function; What is the normalized state after the most probable value is measured?
13	3	Wave function; What are the possible values of repeating measurements for noncommuting operators $\hat{H}$ and $\hat{x}$?

integrity of the participants, deviation from informed consent, studying children under the age of 15, exposure to exceptionally strong stimuli, causing long-term mental harm beyond the risks of everyday life, or risks to the security of the participants. Hence the study did not require an ethics review, according to the guidelines of Finnish Advisory Board on Research Integrity [44]. All the data were pseudonymized before analysis.

C. Data analysis

Latent variable models are statistical models that relate observed variables to unobservable, latent traits. In this study, we used two such models: the Rasch model for the self-efficacy data and latent class analysis for the conceptual knowledge data (cf. Refs. [45,46]).

Rasch analysis converts the raw data (in our case 5-step Likert data) to linear measures. Doing so, the analysis provides an estimate of the probability that a respondent with certain self-efficacy will answer one way or another to an item with a certain agreeableness [47]. In addition, Rasch analysis provides useful tools to investigate the validity and reliability of measurement [48,49].

The latent class analysis (LCA) is a segmentation method for identifying underlying class membership among subjects using categorical variables (cf. Ref. [50]). Hence, the respondents are divided into groups according to patterns in the data, rather than a single metric such as the total score.

All statistical analysis was conducted with R (version 4.1.0) [51] using RStudio [52].

1. Rasch analysis

We used a polytomous extension of the Rasch model called the partial credit model to study the validity of the self-efficacy survey and compute a total self-efficacy score. There are no general sample size guidelines for Rasch analysis, though one lower limit suggested in literature is $N = 50$ [45,53]. As our sample of 272 respondents well exceeds this suggested lower limit, we deemed it appropriate for analysis.

All the analysis related to Rasch was run using the R package *eRm* [54,55], except for computing Spearman's rank correlation [56] that is included in standard R and Cronbach's alpha [57] that was done with the *ltm* package [58] as well as computing and examination of Q_3 statistics that was carried out using the *TAM* package [59].

To investigate the validity and reliability of the measurement, we calculated Spearman's rank correlations for item-rest correlation (in the range 0.49–0.60) [48] and Cronbach's alpha (0.752), indicating an acceptable level of unidimensionality [60]. Examining the residual correlations using Q_3 statistics [61], we saw that they are in the range −0.480–0.043. Even if the minimum value of the residual correlation (between items 2 and 4) is larger than ideal, we deemed it acceptable based on the upper bound of 0.5 for the absolute value of a correlation suggested in literature [62]. Thus we can regard our questionnaire items locally independent.

TABLE IV. Item-fit statistics for the questionnaire items.

Item number	χ^2	p value	Outfit MSQ	Infit MSQ
SE1	193.490	0.997	0.771	0.784
SE2	194.971	0.996	0.777	0.796
SE3	169.527	1.000	0.673	0.636
SE4	245.850	0.580	0.976	0.858

To estimate the data-model fit, we have used the infit and outfit mean-square [63] as well as χ^2 goodness-of-fit statistics shown in Table IV. Examining the χ^2 values, we see that the value for the item 4 is relatively large indicating worse data-model fit. However, the p values do not point to significant deviations from the fit, and the outfit and infit mean-square values are well in the range 0.6–1.4 acceptable for a rating scale survey [48,64]. In addition, we have computed separation reliability [48,65] and person-misfit to estimate the person-fit (i.e., the fit of the response patterns). In our case, the separation reliability is 0.663 which indicates that the items separate between respondents to an acceptable degree [65]. Also the computed person misfit percentage is as low as 2.38%, indicating that the response patterns of only a handful of respondents deviate from the model expectation.

2. Latent class analysis

To examine the knowledge profiles of the students in our sample, we conducted latent class analysis (LCA) for the post-test results. The minimum sample size for LCA depends on the complexity of the survey and model. A lower limit for LCA has been set at $N = 70$ [66], but for simple models, even $N = 30$ may be enough [67]. Hence, our sample ($N = 222$) was deemed sufficient. The analysis was performed using the *poLCA* package [68].

For model selection, we used the parsimony measures Bayesian information criterion (BIC) and Akaike information criterion (AIC), normalized entropy and likelihood ratio chi-square statistic G^2. The model fit indices are shown in Table V. In general, the optimal class solution is the one with the lowest BIC and AIC values, the highest normalized entropy and the lowest G^2 [68,69]. In addition to the fit indices, we evaluated the sizes and average posterior probabilities of the classes, as well as class interpretability. As the other fit indices did not show strong preference for another model, while the AIC was lowest for the model with three groups, we chose this model for further examination.

3. Statistical analysis

To further examine the data after the Rasch and latent class analyses, we used Pearson's χ^2 to examine whether

TABLE V. Fit indices for LCA and class sizes.

No. of classes	BIC	AIC	G^2	Normalized entropy	Class sizes
1	3233.21	3188.97	1074.054	⋯	100%
2	3015.42	2923.55	780.627	0.720	25%, 75%
3	3041.33	2901.82	730.896	0.705	20%, 32%, 48%
4	3093.57	2906.42	707.500	0.699	20%, 24%, 25%, 31%
5	3145.58	2910.79	683.874	0.693	8%, 16%, 19%, 26%, 31%

the observed frequencies deviated from expected frequencies [56]. Where the expected frequencies were less than 5, Fisher's exact test was used [70]. Once a statistically significant difference in samples had been found, the strength of the association (effect size) was determined using Cramer's V [71].

For comparing samples, we first determined whether they were normally distributed with the Shapiro-Wilks test [72,73]. As this was not the case, the Kruskal-Wallis H test was used [74], and if a significant difference was found, the *post hoc* pairwise testing was done with a Wilcoxon rank sum test [74,75], using the Bonferroni correction for the p value for significance [76]. The effect sizes were estimated using Cliff's delta [77]. In addition to standard R, we used packages *effsize* [78] and *rcompanion* [79].

IV. RESULTS AND DISCUSSION

A. The effects of teaching modality

Before studying the effects of prime-time learning and remote teaching on student self-efficacy and conceptual knowledge, we examined the student sample. In terms of self-reported major subjects the sample remained approximately unchanged during the shifts in teaching modality, and the analysis did not show statistically significant differences [$\chi^2\,(4, N = 222) = 2.163$, $p = 0.706$] between the proportions of major subjects. The proportions of different major subjects for 2019–2022 are shown in Fig. 1.

1. The effect of teaching modality on self-efficacy

To study whether the teaching modality had an effect on students' self-efficacy, we compared the self-efficacy scores of the different years. We discovered that there was a statistically significant relationship between the mode of teaching and self-efficacy score with $H(3, N = 272) = 8.168$, $p = 0.043$. The average self-efficacy scores are presented in Table VI. Follow-up tests showed that the distribution of the self-efficacy differed significantly between 2019 and 2020, when the course shifted from active onsite lectures to onsite prime-time learning, and the self-efficacy was significantly higher in 2020 than in 2019 ($W = 861.5$, $p = 0.0082$). The α level for significance is $0.05/6 = 0.0083$ after the Bonferroni correction. The effect size for the difference between the years 2019 and 2020 is small (Cliff's delta -0.31) [80].

Overall, the students self-efficacy remained fairly stable throughout the study and through the teaching reforms. There was a dip in the self-efficacy average scores in 2019, when the spin first approach was first implemented. As the mode of teaching was the same between 2018 and 2019, this effect likely relates to a lack of experiences of mastery or negative feelings experienced by students, rather than a change in interpersonal contacts. Previous research shows that when students encounter new types of study materials, they find it difficult to change their strategy of learning [25].

However, adoption of prime time in 2020 resulted in a significant increase in self-efficacy scores, when comparing the active lectures to onsite prime-time teaching. Hence, the benefits of prime time compensated the negative effects of the new curricular approach. This might be due to the fact that prime time focuses on comprehensive assessment of knowledge and performance. For example, the use of self-assessment supports students' self-efficacy beliefs [19].

TABLE VI. Averages of post-test results (maximum points 13) and self-efficacy measures (ranging between −2.71 and 5.27) by year and LCA group.

	2018	2019	2020	2021–2022	Group A	Group B	Group C
N	50	46	54	122	70	107	45
Mean post-test score	⋯	9.02	8.24	9.09	11.53	8.79	4.93
Mode post-test score	⋯	11	10	11	11	9	6
Standard deviation of post-test scores	⋯	2.534	2.670	2.650	0.896	1.252	1.724
Mean self-efficacy measure	2.20	1.51	2.29	2.15	2.40	2.08	1.45
Mode self-efficacy measure	2.05	1.60	2.60	2.60	2.05	2.60	1.60
Standard deviation of self-efficacy measure	1.690	1.505	1.351	1.478	1.418	1.324	1.726

The self-efficacy beliefs of students remained on the higher level also during remote instruction. The difference in self-efficacy between active onsite lecture and remote prime time (comparing years 2019 and 2021–2022) was not quite statistically significant ($W = 2080,\ \ p = 0.0094$) when accounting for the Bonferroni correction, which is known to be conservative [81].

As the self-efficacy distributions from the onsite prime-time teaching in 2020 and remote prime-time teaching in 2021–2022 did not differ, there is a strong indication that prime time equally supports self-efficacy in a remote setting. These results indicate that the prime-time model, possibly together with the self-evaluation, has enabled students to perform to their own satisfaction.

As the detriments of remote teaching are well established, and the effect is particularly large for beginning physics students [34], it is remarkable how resilient the prime-time model seems in light of our results. In remote teaching, students in general feel less connected to both their peers and the university staff, and their self-efficacy beliefs suffer [35]. Our results however show that in prime-time teaching the personal connection with the teacher can be implemented also through video conference platforms.

It is important to note that we did not measure student well being, and there may be negative effects which our dataset does not show. However, the self-efficacy beliefs were remarkably stable between the implementations of onsite and remote prime-time implementations in our sample.

2. The effect of teaching modality on conceptual learning

To study the effect of the mode of teaching (onsite and remote, active lectures and prime time) to conceptual knowledge post teaching, we compared the results from the conceptual tests based on mode of teaching. The average post-test scores are presented in Table VI.

The analysis showed that the differences in post-test scores between the different modes of teaching were not statistically significant [$H(2) = 5.022,\ p = 0.081$]. Thus, we conclude that the prime-time model had no measurable impact on the learning of quantum mechanics concepts. This is in line with previous results, where the implementation of prime time had a larger effect on student retention than on conceptual learning [12].

In contrast, a recent study surveying physics students in the United States found that regularly working in small groups increased conceptual learning [82]. However, at the University of Helsinki, the physics courses already used collaborative learning strategies prior to prime-time learning, and the majority of students have always attended the collaborative tutored exercise sessions. Hence, it seems that the added value of the prime time was the personal connection to the teacher.

To summarize, during the teaching reforms, the students had a better learning experience, and the learning outcomes remained the same. Importantly, taking away the exam and implementing self-evaluation as part of the grade did not lead to a decrease in conceptual knowledge post teaching.

B. Comparison between LCA groups

Looking purely at the conceptual questions, we wanted to study what kind of learning profiles emerge from the data and to examine the connection between self-efficacy and conceptual learning. As described in Sec. III C 2, we identified three LCA groups of students based on their answers to the abbreviated QMCA post-test: Group A consists of 70 students who performed very well in the post-test, with scores ranging from 10 to 13 out of the maximum 13. The students in group B (107 students) have acquired average or good points (between 6 and 12) in the test, while group C (45 students) contains the students with the lowest scores in the test (between 1 and 8). A summary of the groups' post-test scores are presented in Table VI and the distributions of the scores in Fig. 2.

First we established that there was no statistically significant relationship between teaching modality and LCA group [$\chi^2(4, N = 222) = 7.860,\ p = 0.097$]. In other words, the distribution of students into different LCA groups remained approximately unchanged throughout the study. Instead, we noticed that the self-reported major subject and LCA group were not independent from each other [$\chi^2(4, N = 222) = 10.003,\ \ p = 0.040$]: First, there are more students of theoretical physics than expected in group A. Second, there are less students of theoretical physics and more students of other major subjects (other than experimental or theoretical physics) than expected in group C. The proportions are shown in Fig. 1.

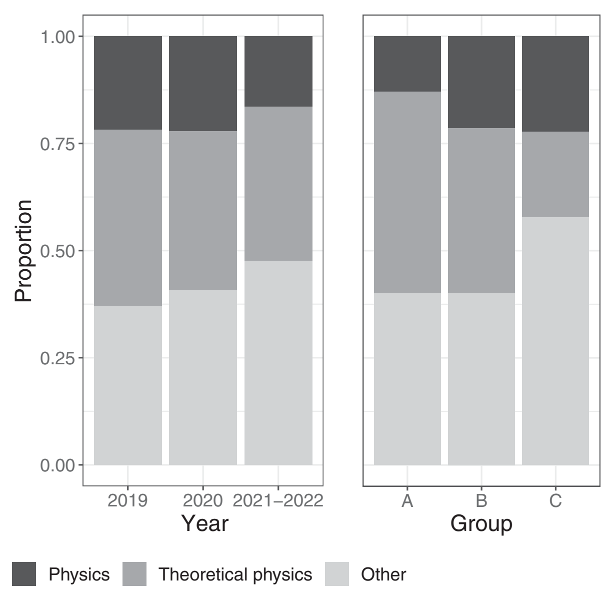

FIG. 1. Proportions of students' self-reported major subjects by year and LCA group.

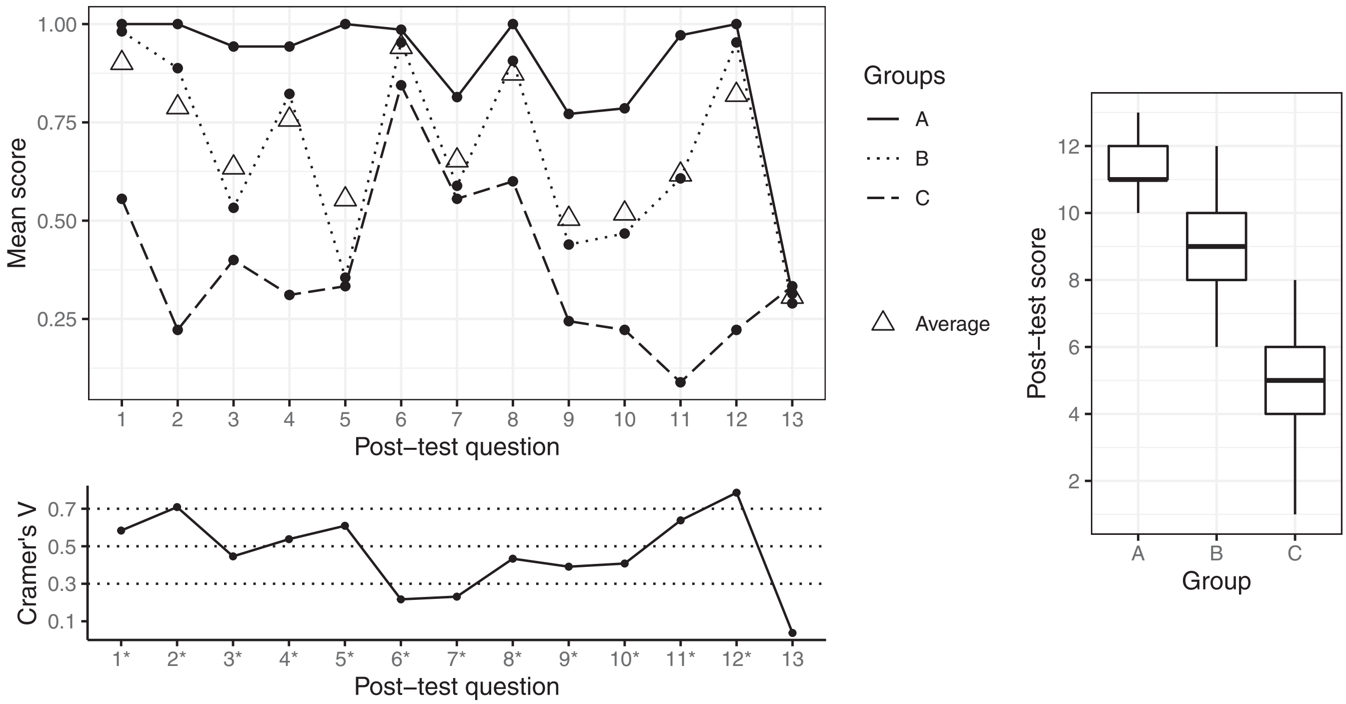

FIG. 2. Item-wise and total post-test scores by LCA group. Top left: item-wise averages for group A (solid line), group B (dotted line), group C (dashed line), and overall average (triangle). Bottom left: Cramer's V (effect size) for intergroup score differences. Significant differences are marked with *. Right: boxplots of post-test scores by group.

1. Learning profiles

To analyze the learning profiles, we evaluated the differences between the LCA groups in terms of conceptual knowledge by comparing the average scores in each question.

As can be seen in Fig. 2, students in group A performed well in almost all questions, but only a few of them managed question 13, which indeed was difficult for all groups and did not differentiate between them ($p > 0.05$; effect size 0.04) [83]. This question deals with repeating measurements for noncommuting operators in the wave function context. The difficulty of this question has been noticed also in previous studies when using the spin first approach [40], and the difficulty may be amplified in an introductory course.

The middle group B is large and their mean scores follow closely the average scores of the whole student sample. Students in group B have stumbled on mostly the same questions as group A, but in addition have had difficulties especially with question 5, which concerns the connection between operators and measurements. This is not surprising, as the misconception that an operator acting on a state constitutes a measurement is known to be confusing for students [84]. This question is seen to differentiate strongly between the groups ($p < 0.05$; effect size 0.61) [83].

In the lowest-performing group C, students have on average answered correctly questions 6 and 7 that differentiate between the groups only slightly ($p < 0.05$; effect sizes 0.22 and 0.23, respectively) [83]. Question 6 asks about a definite spin value for a superposition state and was relatively easy for all groups, while question 7, dealing with a relative phase, was demanding for all groups. Furthermore, the students in group C have on average known what the maximum measurement result is when measuring the spin of a spin-$1/2$ particle (strongly differentiating question 1; $p < 0.05$ and effect size 0.58) and that a property is preserved in subsequent measurements (moderately differentiating question 8; $p < 0.05$ and effect size 0.43) [83].

Interestingly, the question pair 2 and 12, asking about the normalized state of the system after measurement of the most probable value, show the strongest differentiation between groups with $p < 0.05$ and effect sizes 0.71 and 0.79 respectively [83]. This result implies that students who understand the normalized state after measurements can answer questions about it in both spin and wave function contexts.

To summarize, looking at the learning profiles, it seems that even the weaker students grasped the basic concepts of quantum mechanics: they were able to recognize the possible measurement results for spin, they knew that a property is preserved in subsequent measurements, and they had some sense of the concept of relative phase. Stronger students built on that by understanding that the state of a system collapses in measurement and could read probability amplitudes in both the spin and wave function contexts. The most skilled students moreover handled subsequent measurements for noncommuting operators, they knew how to construct a state representation based on the measurement results and recognized the connection between operating and measuring. However, very few of them were able to correctly interpret the infinitely many possible measurement results of the energy of a particle in a box.

2. Self-efficacy in LCA groups

Finally, we studied the relationship between students' group membership and their self-efficacy, and we discovered statistically significant differences between groups [$H(2, N = 222) = 11.636$, $p = 0.003$]. *Post hoc* testing showed that the students in groups A and B rated their self-efficacy significantly better than the students in group C ($W = 2155.5$, $p = 0.001$ and $W = 3014.5$, $p = 0.014$, respectively). The α level for significance is $0.05/3 = 0.017$ with the Bonferroni correction. The difference between the groups A and C is medium (Cliff's Delta 0.37) and that between B and C small (Cliff's Delta 0.25) [80].

Previous research has shown that self-efficacy beliefs mediate learning [4]. In our previous study, we found that the raw scores from the self-efficacy survey were not correlated with students' initial mathematical knowledge but, rather, the students who aimed at theoretical physics reported higher self-efficacy beliefs, even though their mathematical skills were only developing [43]. Now we see that theoretical physics students are overrepresented in group A, which leads us to conclude that their self-efficacy beliefs are in line with their conceptual knowledge post teaching.

These results offer an interesting view into the role of self-efficacy and physics identity. Physics self-efficacy and physics identity are largely aligned [17], and indeed some view self-efficacy as a subcomponent of physics identity [18]. At least in Finland, the question of physics identity has been problematic, with theoretical physics being more highly valued by physicists [24]. Our results show that students who intend to major in theoretical physics have higher self-efficacy, and as is typical in the literature, higher self-efficacy correlates with stronger conceptual knowledge. From a snapshot of our students, we cannot say how students' self-efficacy develops and why students choose their major subject. Clearly more research is needed on this topic.

There is a well-established link between self-efficacy and gender [21,22], as well as between gender and physics identity [17,85], and theoretical physics is one of the subfields of physics with the least women. On the other hand, some physicists perceive theoretical physics as the only real physics and both male and female physicists of other subfields may struggle with identifying as physicists [24]. As the interplay between gender, interest, and identity is very complex, physics should not be viewed as a single field when studying physics identity.

V. CONCLUSIONS

We surveyed students' self-efficacy and conceptual knowledge post teaching through a transition from the position first to spin first instructional approach and a shift from active lectures to prime-time learning. These changes preceded the COVID-19 pandemic, which provided us the opportunity to study the effects of prime-time learning during remote teaching.

We saw that students' self-efficacy beliefs were negatively affected by the transition to spin first teaching, but the introduction of prime time improved self-efficacy. Contrary to findings from other studies, we did not see a decline in self-efficacy during the remote instruction.

Post-test scores were satisfactory throughout the study, and we did not see statistically significant changes in them during the changes in teaching modality. Thus it seems that the prime-time model did not significantly improve conceptual learning. Instead, the model proved its resilience in the transition to remote teaching, as students' post-test scores did not suffer.

To study QM learning profiles, students were distributed into groups based on their post-test answers. We saw that the distribution of students into different groups was not affected by mode of teaching.

The learning profiles show that knowledge of quantum mechanics is built up from the basic understanding about a single measurement in a two-state system and repeated measurements of the same observable. This understanding is required for grasping the idea of repeated measurements for different—both commuting and noncommuting—operators. Most of our students, who were taught in the spin first approach, gained also the basic understanding of the wave function representation but still struggled with the infinitely many possible measurement results of the energy of a particle in a box.

Students in the two highest-performing groups had significantly higher self-efficacy scores than the students in the third group. Students of theoretical physics were over-represented in the highest-performing group with the highest average self-efficacy score, while the students with a major subject other than experimental or theoretical physics were over-represented in the lowest-performing group with the lowest average self-efficacy score. This finding is in line with the previous studies showing that, on the one hand, self-efficacy mediates learning and, on the other hand, that physics courses may not support self-efficacy of students who are not perceived as typical physics majors. Our results show that the interplay between physics interest, choice of a major subject, and competency beliefs is highly complex, and more research is needed on the relationship between, for example, gender and choice of physics subfield.

ACKNOWLEDGMENTS

We want to express our sincere gratitude to Associate Professor David Weir for help with the translation of the conceptual test and for careful reading of the manuscript. We thank Professor Homeyra Sadaghiani for discussions of the QMCA questions. The support of Associate Professor Iasonas Lamprianou for the Rasch, Dr. Tommi Kokkonen and Dr. Neea Palojärvi for the LCA, and Dr. Jokke Häsä for the statistical analysis is gratefully acknowledged. Moreover, we wish to thank M.Sc. Tapio Rasa for help with structuring parts of the manuscript. The helpful reviews of the two anonymous reviewers are gratefully acknowledged. The work of E. P. was supported by The Magnus Ehrnrooth foundation. Open access funded by Helsinki University Library.

[1] M. V. Be, E. K. Henriksen, T. Lyons, and C. Schreiner, Participation in science and technology: Young peoples achievement related choices in late modern societies, Studies Sci. Educ. **47**, 37 (2011).

[2] A. Bandura, Self-efficacy, in *Encyclopedia of Human Behavior*, edited by V. S. Ramachaudran (Academic Press, New York, 1994), pp. 71–81.

[3] V. Sawtelle, E. Brewe, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, J. Res. Sci. Teach. **49**, 1096 (2012).

[4] B. J. Zimmerman, Self-efficacy: An essential motive to learn, Contemp. Educ. Psychol. **25**, 82 (2000).

[5] A. Johansson, S. Andersson, M. Salminen-Karlsson, and M. Elmgren, Shut up and calculate: The available discursive positions in quantum physics courses, Cultural Studies Sci. Educ. **13**, 205 (2018).

[6] H. Sadaghiani, Spin first vs. position first instructional approaches to teaching introductory quantum mechanics, in *Proceedings of PER Conf. 2016, Sacramento, CA* (AIP, New York, 2016), 10.1119/perc.2016.pr.068.

[7] G. Zhu and C. Singh, Improving students understanding of quantum mechanics via the Sterngerlach experiment, Am. J. Phys. **79**, 499 (2011).

[8] C. A. Manogue, E. Gire, D. McIntyre, and J. Tate, Representations for a spins-first approach to quantum mechanics, AIP Conf. Proc. **1413**, 55 (2012).

[9] H. R. Sadaghiani and J. Munteanu, Spin first instructional approach to teaching quantum mechanics in sophomore level modern physics courses, in *Proceedings of PER Conf. 2015, College Park, MD* (AIP, New York, 2015), 10.1119/perc.2015.pr.067.

[10] M. Prince, Does active learning work? A review of the research, J. Engin. Educ. **93**, 223 (2004).

[11] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).

[12] P. Koskinen, J. Lämsä, J. Maunuksela, R. Hämäläinen, and J. Viiri, Primetime learning: Collaborative and technology-enhanced studying with genuine teacher presence, Int. J. STEM Educ. **5**, 1 (2018).

[13] L. Heikkinen, I. Kontro, P. Koskinen, J. Maunuksela, M. Saarelainen, and K. Tuominen, Fysiikan opetuksen kehitystyö auttoi vastaamaan korona-ajan haasteisiin, Arkhimedes **1**, 16 (2021), https://journal.fi/arkhimedes/article/view/113524/67008.

[14] D. H. Schunk and F. Pajares, The development of academic self-efficacy, *Development of Achievement Motivation* (Elsevier, New York, 2002), pp. 15–31.

[15] S. L. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, J. Res. Sci. Teach. **43**, 485 (2006).

[16] R. W. Lent, H.-B. Sheu, D. Singley, J. A. Schmidt, L. C. Schmidt, and C. S. Gloster, Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students, J. Vocat. Behav. **73**, 328 (2008).

[17] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, Phys. Rev. Phys. Educ. Res. **15**, 020119 (2019).

[18] Z. Hazari, G. Sonnert, P. M. Sadler, and M.-C. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47**, 978 (2010).

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

[20] R. Dou, E. Brewe, J. P. Zwolak, G. Potvin, E. A. Williams, and L. H. Kramer, Beyond performance metrics: Examining a decrease in students' physics self-efficacy through a social networks lens, Phys. Rev. Phys. Educ. Res. **12**, 020124 (2016).

[21] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 020105 (2016).

[22] E. M. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, Phys. Rev. Phys. Educ. Res. **14**, 020123 (2018).

[23] 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).

[24] J. Vainio, Hegemony, contradiction, and gender in the context of Finnish university physics, Ph.D. thesis, University of Helsinki, 2012.

[25] M. V. Be, E. K. Henriksen, and C. Angell, Actual versus implied physics students: How students from traditional physics classrooms related to an innovative approach to quantum physics, Sci. Educ. **102**, 649 (2018).

[26] E. Gire and E. Price, Structural features of algebraic quantum notations, Phys. Rev. ST Phys. Educ. Res. **11**, 020109 (2015).

[27] B. P. Schermerhorn, G. Passante, H. Sadaghiani, and S. J. Pollock, Exploring student preferences when calculating expectation values using a computational features framework, Phys. Rev. Phys. Educ. Res. **15**, 020144 (2019).

[28] T. Rasa, E. Palmgren, and A. Laherto, Futurizing science education: Students' experiences from a course on futures thinking and quantum computing, Instr. Sci. **50**, 425 (2022).

[29] Opetushallitus, Fysiikan tukimateriaalia (2021), https://www.oph.fi/sites/default/files/documents/lops2019_fy_0.pdf.

[30] J. Vettenranta, J. Välijärvi, A. Ahonen, J. Hautamäki, J. Hiltunen, K. Leino, S. Lähteinen, K. Nissinen, V. Nissinen, E. Puhakka, J. Rautopuro, and M.-P. Vainikainen, PISA 15 Ensituloksia. Huipulla pudotuksesta huolimatta, Tech. Rep. 41 (Opetus- ja kulttuuriministeriö, 2016).

[31] J. Lavonen and S. Laaksonen, Context of teaching and learning school science in Finland: Reflections on PISA 2006 results, J. Res. Sci. Teach. **46**, 922 (2009).

[32] D. L. Vargas, A. M. Bridgeman, D. R. Schmidt, P. B. Kohl, B. R. Wilcox, and L. D. Carr, Correlation between student collaboration network centrality and academic performance, Phys. Rev. Phys. Educ. Res. **14**, 020112 (2018).

[33] C. Stadtfeld, A. Vörös, T. Elmer, Z. Boda, and I. J. Raabe, Integration in emerging social networks explains academic failure and success, Proc. Natl. Acad. Sci. U.S.A. **116**, 792 (2019).

[34] P. Klein, L. Ivanjek, M. N. Dahlkemper, K. Jeličić, M.-A. Geyer, S. Küchemann, and A. Susac, Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories, Phys. Rev. Phys. Educ. Res. **17**, 010117 (2021).

[35] I. Marzoli, A. Colantonio, C. Fazio, M. Giliberti, U. Scotti di Uccio, and I. Testa, Effects of emergency remote instruction during the COVID-19 pandemic on university physics students in italy, Phys. Rev. Phys. Educ. Res. **17**, 020130 (2021).

[36] J. Häsä, J. Rämö, and V. Virtanen, Evaluating students' self-assessment in large classes, in *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education*, edited by U. Jankvist, M. van den Heuvel-Panhuizen, and M. Veldhuis (Freudenthal Group & Freudenthal Institute, Utrecht University, Netherlands, 2019), pp. 4083–4090.

[37] R. Harris, *Modern Physics* (Pearson/Addison Wesley, London, 2008).

[38] D. H. McIntyre, C. A. Manogue, and J. Tate, *Quantum Mechanics: A Paradigms Approach* (Pearson, London, 2014).

[39] J. M. Bailey, D. Lombardi, J. R. Cordova, and G. M. Sinatra, Meeting students halfway: Increasing self-efficacy and promoting knowledge change in astronomy, Phys. Rev. Phys. Educ. Res. **13**, 020140 (2017).

[40] J. Wells, H. Sadaghiani, B. P. Schermerhorn, S. Pollock, and G. Passante, Deeper look at question categories, concepts, and context covered: Modified module analysis of quantum mechanics concept assessment, Phys. Rev. Phys. Educ. Res. **17**, 020113 (2021).

[41] H. R. Sadaghiani and S. J. Pollock, Quantum mechanics concept assessment: Development and validation study, Phys. Rev. ST Phys. Educ. Res. **11**, 010110 (2015).

[42] S. Pollock, H. Sadaghiani, A. Quaal, and G. Passante, Designing, validating, and contrasting conceptual quantum mechanics questions for spin states and spatial wave functions , in *Proceedings of PER Conf. 2018, Washington, DC* (AIP, New York, 2018), 10.1119/perc.2018.pr.Pollock.

[43] I. Kontro and E. Palmgren, The role of mathematics and self-efficacy in learning quantum mechanics, J. Phys. Conf. Ser. **1512**, 012001 (2020).

[44] Finnish National Board on Research Integrity, Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review (2009), https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf.

[45] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, Rasch analysis in physics education research: Why measurement matters, Phys. Rev. Phys. Educ. Res. **15**, 020111 (2019).

[46] M. Hickendorff, P. A. Edelsbrunner, J. McMullen, M. Schneider, and K. Trezise, Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis, Learning Indiv. Dif. **66**, 4 (2018).

[47] A. R. Conway and K. Kovacs, *Individual Differences in Intelligence and Working Memory: A Review of Latent Variable Models* (Academic Press, New York, 2013), pp. 233–270.

[48] I. Lamprianou, *Applying the Rasch Model in Social Sciences Using R and BlueSky Statistics* (Routledge, London, 2019).

[49] W. J. Boone, Rasch analysis for instrument development: Why, when, and how?, CBE Life Sci. Educ. **15**, 1 (2016).

[50] K. E. Masyn, Latent class analysis and finite mixture modeling, in *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis*, edited by T. D. Little (Oxford University Press, New York, 2013), Chap. 25, pp. 551–611.

[51] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020).

[52] RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio (PBC, Boston, MA, 2020).

[53] J. M. Linacre, *Many-Facet Rasch Measurement* (MESA Press, San Diego, CA, 1994).

[54] P. Mair, R. Hatzinger, and M. J. Maier, eRm: Extended Rasch Modeling (2021), R package version 1.0-2, https://cran.r-project.org/package=eRm.
[55] P. Mair and R. Hatzinger, Extended Rasch modeling: The eRm package for the application of IRT models in R, J. Stat. Softw. **20**, 1 (2007).
[56] D. G. Rees, *Essential Statistics* (Springer, New York, 2000).
[57] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16**, 297 (1951).
[58] D. Rizopoulos, ltm: An R package for latent variable modelling and item response theory analyses, J. Stat. Softw. **17**, 1 (2006).
[59] A. Robitzsch, T. Kiefer, and M. Wu, TAM: Test Analysis Modules (2021), R package version 3.7–16, https://CRAN.R-project.org/package=TAM.
[60] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference*, 11.0 update (Allyn & Bacon, Boston, 2003).
[61] W. M. Yen, Effects of local item dependence on the fit and equating performance of the three-parameter logistic model, Appl. Psychol. Meas. **8**, 125 (1984).
[62] M. Davidson, J. Keating, and S. Eyres, A low back-specific version of the SF-36 physical functioning scale, Spine **29**, 586 (2004).
[63] B. D. Wright and M. H. Stone, *Best Test Design* (MESA Press, San Diego, CA, 1979).
[64] B. D. Wright and J. M. Linacre, Reasonable mean-square fit values, Rasch Meas. Trans. **8**, 370371 (1994), https://www.rasch.org/rmt/rmt83b.htm.
[65] B. D. Wright and M. H. Stone, *Measurement Essentials* (Wide Range Inc., Grimsby, ON, Canada, 1999).
[66] I. C. Wurpts and C. Geiser, Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study, Front. Psychol. **5**, 1 (2014).
[67] G. H. Lubke, Latent variable mixture modeling, in *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, edited by G. R. Hancock and R. O. Mueller (Routledge, New York, 2019), Chap. 15, pp. 202–213.
[68] D. A. Linzer and J. B. Lewis, poLCA: An R package for polytomous variable latent class analysis, J. Stat. Softw. **42**, 129 (2011).
[69] D. Nagin, *Group-Based Modeling of Development* (Harvard University Press, Cambridge, MA, 2009).
[70] R. A. Fisher, *Statistical Methods for Research Workers* (Oliver & Boyd, Edinburgh, UK, 1970).
[71] H. Cramr, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
[72] J. P. Royston, An extension of Shapiro and Wilk's W test for normality to large samples, J. R. Stat. Soc.. Series C (Applied Statistics) **31**, 115 (1982).
[73] S. S. Shapiro and M. B. Wilk, An analysis of variance test for normality (complete samples), Biometrika **52**, 591 (1965).
[74] M. Hollander, *Nonparametric Statistical Methods* (Wiley, New York, 2013).
[75] D. F. Bauer, Constructing confidence sets using rank statistics, J. Am. Stat. Assoc. **67**, 687 (1972).
[76] J. P. Shaffer, Multiple hypothesis testing, Ann. Rev. Psychol. **46**, 561 (1995).
[77] N. Cliff, *Ordinal Measurement in the Behavioral Sciences* (Routledge, London, 1996).
[78] M. Torchiano, Effsize—a Package for Efficient Effect Size Computation, 10.5281/zenodo.1480624 (2020), R package version 0.8.1.
[79] S. Mangiafico, Rcompanion: Functions to Support Extension Education Program Evaluation (2021), R package version 2.4.1, https://CRAN.R-project.org/package=rcompanion.
[80] J. Romano, J. D. Kromrey, J. Coraggio, and J. Skowronek, Appropriate statistics for ordinal level data: Should we really be using t test and Cohen's d for evaluating group differences on the NSSE and other surveys?, in *Proceedings of the Annual Meeting of the Florida Association of Institutional Research, Cocoa Beach, Florida* (FAIR, Florida, 2006), pp. 1–3.
[81] P. Sedgwick, Multiple significance tests: the Bonferroni correction, Br. Med. J. **344**, 509 (2012).
[82] M. Rodriguez and G. Potvin, Frequent small group interactions improve student learning gains in physics: Results from a nationally representative pre-post study of four-year colleges, Phys. Rev. Phys. Educ. Res. **17**, 020131 (2021).
[83] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1988).
[84] E. Gire and C. Manogue, Making sense of quantum operators, eigenstates and quantum measurements, AIP Conf. Proc. **1413**, 195 (2012).
[85] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way, Phys. Rev. Phys. Educ. Res. **15**, 020148 (2019).