

Assessing changes in physics students' preparedness: Mathematics and physics prior knowledge of German first-semester students in 2013 and 2023

Dennys Gahrmann^{1,*}, Irene Neumann², and Andreas Borowski¹

¹*Physics Education, Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany*

²*Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany*



(Received 2 September 2025; accepted 3 December 2025; published 28 January 2026)

Insufficient prior knowledge in mathematics and physics is seen as a key factor for high attrition rates in physics degree programs. To investigate trends in students' academic preparedness, this study examines how the mathematics and physics prior knowledge of first-year physics students in Germany has changed between 2013 and 2023. Using data from standardized diagnostic tests administered at multiple universities in both cohorts, we compared performance across cohorts and institutions. Results indicate small but meaningful increases in average performance from 2013 to 2023: mathematics knowledge improved slightly (Cohen's $d = 0.10$), and physics knowledge even more (Cohen's $d = 0.21$). Despite this improvement, students' average performance remains only modest—67% correct in mathematics and only 38% correct in physics. Moreover, our findings indicate a growing heterogeneity in students' preparation: while mathematics and physics performance improved at some universities, mathematics performance declined at others. Overall, the results underscore the importance of continued monitoring of incoming students' prior knowledge and support the use of diagnostic tools to inform targeted instructional strategies transition to university-level physics.

DOI: [10.1103/vkyq-fmcv](https://doi.org/10.1103/vkyq-fmcv)

I. INTRODUCTION

Science, technology, engineering, and mathematics (STEM) degree programs face low completion rates worldwide [1,2], and, for example, in the United States, six-year degree completion rates in STEM remain below 40% [3]. Student retention is therefore an important issue for higher education research. International research consistently identifies study preparedness as a key determinant of academic success and student retention in university-level STEM programs, particularly in physics [4–6]. Although the relevance of study preparedness is observed across countries, the responsibility for addressing and supporting it lies primarily at the national level. In the absence of coordinated global strategies, it becomes all the more important for individual countries to ensure that first-year physics students are adequately prepared to meet the specific academic challenges they will face [7,8].

Among the dimensions of study preparedness— affective, personal, cognitive, and content-related (see Refs. [5,9–11])—the content-related dimension, particularly prior knowledge, plays a major role in student success

and retention [4,12–14]. Numerous studies have demonstrated that students' prior knowledge in mathematics and physics is a strong predictor of their performance in university physics. These studies typically use regression or other predictive models to show that students with higher levels of prior knowledge tend to perform better in physics exams. These findings consistently highlight the central role of prior knowledge in explaining academic success in physics programs, e.g., [6,7,15,16]. Research suggests that ancillary prior knowledge, such as mathematics prior knowledge in a university physics study program, is even more important for academic success than physics prior knowledge [4]. Despite its high relevance, prior research has typically aggregated cohorts across years, thereby implicitly assuming that students' mathematics and physics knowledge remains stable over time [8,17,18]. This assumption overlooks the possibility that prior knowledge may change across cohorts. As general educational contexts evolve, students' knowledge may shift over time. For example, recent large-scale assessments such as the 2022 NAEP study in the United States [19] and the 2022 PISA study in Germany [20] indicate changes in students' mathematical skills. It is therefore reasonable to assume that the prior knowledge relevant to studying university-level physics also changes over time. This gap— changes in prior knowledge for cohorts of different years—is important to analyze because overlooking such changes may

*Contact author: Dennys.gahrmann@uni-potsdam.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

lead to misinterpretations of student preparedness, hinder the effectiveness of study program design, and lead to drop out.

Addressing this issue, Buschhüter *et al.* conducted a large-scale assessment comparing cohorts from 1978 and 2013 in Germany, revealing significant changes in physics prior knowledge—with medium effect sizes [7]—and statistically significant but negligible changes in mathematics prior knowledge [15]. Building on the study by Buschhüter *et al.*, we compared *physics* prior knowledge between cohorts from 2013 and 2023 [21] and found statistically significant differences. Yet, in contrast to the findings by Buschhüter *et al.* [7], the effect was negligible. So far, our analyses did not include students' mathematics prior knowledge, although it is known to play a key role in physics learning (see also Ref. [22]).

Building on the above studies, the present study examines whether and how mathematics prior knowledge has evolved over the same period and whether these changes mirror trends in physics prior knowledge by comparing first-year physics students in Germany from 2013 and 2023. By explicitly examining the differences between two cohorts a decade apart, we respond to earlier calls and findings [7,15,19] and examine whether the assumptions of stability in prior knowledge still hold in today's context. In addition, we examine differences between changes in mathematics prior knowledge and physics prior knowledge reported by Gahrman *et al.* [21] to provide a comprehensive picture of student readiness. Finally, we provide an item-level analysis of both mathematics and physics items with the aim of identifying specific areas of cohort differences, thereby extending our previous findings [21].

II. THEORY AND BACKGROUND

A. Study preparedness

Study preparedness can be defined as “a set of personal characteristics whose successful development within an institutional context enables and makes student success likely, either in general, within a group of study programs, or in a specific study program” [23]. Student success, mostly relevant for retention rates and persistence of students, is an outcome of study preparedness (see, e.g., Ref. [10]). Student success is often categorized into three domains: objective (measurable outcomes, such as competence development; see Refs. [24,25]), certified (formal academic achievements, such as grades, completed modules, and academic degrees; see Ref. [7]), and subjective criteria (students' personal perceptions of success, including, e.g., students' satisfaction with their academic performance, see Ref. [26] [23]).

The above definition of study preparedness emphasizes the importance of individual dispositions that influence students' ability to adapt and succeed in the academic environment. It highlights that study preparedness is about possessing dispositions that support student

success. Theoretical and empirical studies describe four (partly overlapping) dimensions of study preparedness: the cognitive, affective, personal, and content-related dimensions [5,9–11].

The cognitive dimension includes intellectual abilities, such as problem solving, critical thinking, and general knowledge. In physics, these skills are essential, for example, for understanding (abstract) concepts and solving (complex) problems, e.g., [27].

The affective dimension involves emotional and motivational factors that influence learning, including attitudes toward challenges, motivation to engage with content, and self-efficacy. These are particularly important in physics, e.g., where demanding material can impact students' motivation and persistence, e.g., [26,28].

The personal dimension covers self-regulation, resilience, time management, and social skills, among others. These help students manage academic demands and learn independently. For example, resilience allows students to cope with setbacks, while social skills support positive interactions with peers and instructors, e.g., [29,30].

The content-related dimension refers to students' prior knowledge and content-related skills. In physics, this includes mathematics prior knowledge as ancillary knowledge and physics prior knowledge as domain-specific knowledge. While these aspects are subject-specific, they are closely related to the cognitive dimension, as they reflect the application of intellectual abilities—such as problem solving and conceptual understanding—but the focus lies more *in the specific disciplinary contexts*.

Prior knowledge, particularly in mathematics and physics, has consistently emerged as a significant predictor of academic performance in physics courses [4,6,7,12,15,16]. Research indicates that mathematics prior knowledge often exerts a stronger influence on performance in physics examinations than prior knowledge in physics itself [4].

Summarizing, prior knowledge as an aspect of the content-related domain of student preparedness is particularly predictive for student success at the beginning of a university physics study program. For more details, see Sec. II C.

B. Changes in general knowledge

Public discourse, see, for example, Thomas and Brown [31], increasingly recognizes that general knowledge changes and evolves over time, both at the societal level and in the educational system. Technological developments, ranging from the widespread use of computers and the emergence of artificial intelligence, have reshaped what is considered general knowledge and also how knowledge is acquired and used [32]. This not only relates to the knowledge of students they bring to school but also to the knowledge and skills students develop through formal education [33].

Evidence for changes in general knowledge has been observed in national studies, e.g., in the United States, such as the National Assessment of Educational Progress (NAEP [34]), which assessed mathematics and reading competencies among students in grades 4 and 8. The 2022 NAEP study showed that while overall mathematics and reading scores in grade 8 remained stable, performance varied by content area—for example, algebra scores dropped from 2022 to 2024, while geometry scores improved [19].

In a similar vein, the 2022 PISA results for Germany indicated a notable decline in the performance of 15-year-olds across all domains—mathematics, reading, and science—compared to 2018. More specifically, the average scores of German students in mathematics decreased from 500 in 2018 to 475 in 2022, and in reading from 498 to 480. These declines are roughly equivalent in magnitude to the typical learning gains students achieve over the course of a full academic year around the age of 15 [35]. The decline from 2018 to 2022 underscores a reversal of the previously positive trend that had begun earlier [20]. It is important to mention that the standard deviation in PISA, based on the whole OECD sample, is 100 [36]. In science, the average score declined from 503 to 492. These reductions of 25, 18, and 11 points, respectively, represent the largest drops recorded for Germany in PISA's history, aligning with the 2023 OECD average. Germany exhibited a slightly higher average score in science compared to the OECD average. The proportion of students achieving only very basic competencies (below PISA proficiency level 2) has increased substantially from 2018 to 2022: approximately 30% in mathematics, 26% in reading, and 23% in science. In contrast, the share of high-performing students has declined, especially in mathematics and reading [20]. These shifts represent a reversal in Germany of the progress made after the initial PISA cycle in 2000 and point to a broader deterioration in student competencies rather than random variation.

The 2023 Survey of Adult Skills conducted by OECD revealed similar patterns of general knowledge internationally when assessing the literacy and numeracy of adults across different age groups, see Ref. [37]. In Germany, the average proficiency in literacy (266) and numeracy (273) of adults of all ages in the study (16–65) is above the OECD average (literacy: 260 and numeracy: 263). Even though the average proficiency in literacy and numeracy has remained stable over the past decade (changes in the past decades: literacy 0.00, numeracy 4.80), differences are visible across age groups: For example, young adults born between 1989 and 1996 scored 12 points higher in literacy in 2022–2023 (as 27- to 34-year-olds) than in 2011–2012 (as 16- to 23-year-olds), indicating a significant improvement. The gap between the highest and lowest performing adults widened in both literacy and numeracy between 2011–2012

and 2022–2023. While the proportion of low-achieving adults remained stable, the proportion of particularly high-achieving adults increased [37]. No information is available about differences regarding specific content areas, but it is observable that the general knowledge of individuals entering university—typically first-semester students—appears to shift over time. However, rather than indicating a general increase or decrease, these changes reflect a shift in the composition or nature of prior knowledge [37].

C. Project background and changes in physics-specific prior knowledge

This paper is part of a larger project that investigates the development of mathematics and physics prior knowledge among German first-semester physics students since 1978. Buschhüter *et al.* analyzed mathematics and physics prior knowledge between the 1978 and 2013 cohorts. More recently, Gahrmann *et al.* [21] examined changes in physics prior knowledge by comparing the 2013 cohort with a 2023 cohort, using a new, validated version of the instrument used in the 2013 study [7,15].

The physics findings reported by Buschhüter *et al.* (Cohen's $d = 0.53$; significant decline between 1978 and 2013) contrast with the more recent findings of Gahrmann *et al.*, which show a statistically significant better performance with a small effect size (Cohen's $d = 0.21$) from 2013 to 2023. This suggests a stabilization of physics prior knowledge. Additionally, Buschhüter *et al.* found a statistically significant change in mathematics prior knowledge with a small effect size (Cohen's $d = 0.2$) between 1978 and 2013. An analysis of mathematics prior knowledge changes from 2013 to 2023 is missing and constitutes the central aim of this paper.

To investigate the mathematics and physics prior knowledge in 2023, the 2013 study entrance test (hereafter abbreviated SET, followed by the year of administration, e.g., SET 2013) used by Buschhüter *et al.* [7,15] was carefully revised. The following section provides a detailed description of SET 2023 (in comparison to SET 2013), covering several key aspects: general information such as the number of items, content validity related to item selection, measurement invariance, and internal consistency.

The SET 2023 incorporates items from two primary sources: (i) $N = 17$ mathematics and $N = 17$ physics anchor items, which are retained from the SET 2013 version to ensure continuity and enable direct comparisons. Additionally, (ii) $N = 20$ mathematics and $N = 15$ physics items adapted from established instruments used in the assessment of physics and mathematics at the beginning of physics studies, e.g., the Lawson Classroom Test [38,39] and the TUG-K [40,41], are included. To compare the 2013 and 2023 cohorts, the anchor items are used, while the additional items are excluded from this paper, as they are not psychometrically comparable across cohorts and thus do not support a valid cross-cohort comparison.

TABLE I. Criteria for measurement invariance between mathematical items of SET 2013 and SET 2023. Note that the differences in CFI and TLI for a model are calculated between a model and the model above. Criteria indicating measurement invariance are $\text{RMSEA} < 0.08$, $\text{CFI} > 0.9$, $\text{TLI} > 0.9$ and for successive models additionally $\Delta\text{CFI} < 0.01$ and $\Delta\text{TLI} < 0.05$, see Hälsig [48]. RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker-Lewis index.

Model	χ^2 test			RMSEA	CFI	TLI	ΔCFI	ΔTLI
	d.o.f.	Test power	p value					
Configural invariance model	238	1235.610	< 0.001	0.044	0.929	0.919		
Metric invariance model	254	1291.454	< 0.001	0.043	0.926	0.921	0.003	0.002
Scalar invariance model	270	1562.894	< 0.001	0.047	0.908	0.907	0.02	0.01
Partial scalar invariance model	262	1450.269	< 0.001	0.046	0.916	0.912	0.01	0.01*

Both SET 2013 and SET 2023 include questions on demographics, such as GPA (grade point average) or participation in preparatory courses before the test was administered. It is important to note that the GPA is not a subject-specific predictor but a general indicator of school performance. It reflects both cognitive and personal aspects of study preparedness, as it includes not only subject grades but also evaluations of social behavior. Furthermore, it is important to mention that in Germany, no SAT (Scholastic Assessment Test) or ACT (American College Testing) is conducted as in the United States, but only the GPA is used as a prerequisite for most university study programs.

The content validity of the items was ensured through three key empirical investigations. First, in 1976, the original item pool was developed based on university physics teachers' perceptions of what knowledge is important for a physics degree program. It focused on content deemed essential for understanding and following physics lectures and was assumed to be necessary for success in examinations [42]. Second, a follow-up investigation by Buschhüter *et al.* in 2013 reaffirmed the relevance of these tasks from the perspective of university physics teachers [43]. Third, two nationwide online surveys (one on the mathematics and one on the physics items) were conducted in 2023. In the surveys, $N = 75$ and $N = 22$ university physics teachers from institutions across Germany rated items from the SET 2013 and selected items from established instruments. Only those items that were in line with current national educational standards were rated. Results from the surveys were used to identify the most suitable items for inclusion in SET 2023, thus ensuring strong content validity based on expert judgment from the university physics teaching community [44].

In summary, the revision of SET 2023 was carefully designed to align the instrument with contemporary curricular standards while preserving its historical roots and the viewpoints of university physics instructors. The validation in both 2013 and 2023 highlights the sustained importance of the chosen items and content domains. Most notably, the items used for comparing the two cohorts were explicitly confirmed by university physics teachers as relevant in 2013 as well as in 2023. This strong

endorsement underscores their continued significance for assessing foundational knowledge at the transition to university-level physics.

The SET 2023 was designed to capture the broad construct of content-related prior knowledge for a university physics study program. In the following section, we focus on the mathematics items and thus on the underlying construct of subject-relevant mathematics prior knowledge for a university physics study program. The relevant mathematics prior knowledge is measured through items from different content areas. We want to differentiate three content areas, which are in line with the parts of relevant mathematics in a physics program described in other studies [45]: basics (such as fractions, equations), differentiations (e.g., deriving a polynomial function), integrals (e.g., integrating polynomial functions), and simple vector calculus (e.g., adding two vectors graphically and symbolically).

Beyond the described content validity, measurement invariance for all mathematics items across both SETs was examined using the statistics software R and the R package LAVAAN [46,47], such that the prerequisites were fulfilled to compare both cohorts. Measurement invariance ensures that the test measures the same construct in the same way across groups, so that observed differences in scores reflect true differences in the underlying ability rather than artifacts of the measurement instrument. Following the procedure by Hälsig [48], a stepwise approach was applied to test measurement invariance regarding the $N = 17$ mathematics items used in 2013 and 2023: First, configural invariance (i.e., the basic factor structure is the same across both cohorts) was tested using a baseline model that includes all items in mathematics. This model achieved a very good fit within acceptable parameter ranges, see Table I. Next, metric invariance (factor loadings are the same, i.e., construct has the same meaning across groups) was assessed by constraining the factor loadings. The resulting model showed a very good fit, with small changes in model parameters compared to the baseline model, see Table I. Finally, partial scalar invariance (intercepts are equal, i.e., latent means across groups can be compared) was tested by constraining all factor loadings

and constraining $N = 9$ intercepts. This model also demonstrated a very good fit, with small changes in model parameters relative to the metric invariance model. Total scalar invariance (constraining all factor loadings and all intercepts) is not necessary to compare both tests, even though it was also tested as recommended in Ref. [49]. The model parameters are presented in Table I for mathematics. To evaluate the models, χ^2 tests were conducted, which assess the discrepancy between the observed and expected data under the specified model. As indicators of model fit, root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) were used, following the recommendations of Halsig. RMSEA quantifies how well a model, with optimally estimated parameters, fits the population covariance matrix. CFI and TLI compare the fit of the specified model to a baseline model, with TLI additionally adjusting for model complexity.

The results indicate that the partial scalar invariance model fits the data well, supporting the assumption of measurement invariance between SET 2013 and SET 2023. With the demonstrated partial scalar invariance, a comparison of means between both cohorts is possible.

In addition to the measurement invariance analysis, the internal consistency of the mathematics and physics tests was evaluated for both SET 2013 and SET 2023, for mathematics and physics each (focusing again only on the $N = 17$ mathematics and $N = 17$ physics items). The reliability coefficients (Cronbach's alpha α) of the mathematics part (2013: $\alpha = 0.83$; 95% CI = [0.82, 0.84], 2023: $\alpha = 0.83$; 95% CI = [0.82, 0.84]) and the physics part and their confidence intervals (CI) (2013: $\alpha = 0.72$; 95% CI = [0.70, 0.74], 2023: $\alpha = 0.74$; 95% CI = [0.72, 0.76]) indicate good to very good internal consistency, see Ref. [50]. These results suggest that all items included in both mathematics and physics measure the underlying construct consistently across both test administrations. Since the items were selected to reflect the knowledge needed at the start of a university physics study program, the underlying construct of the mathematics items is mathematics knowledge relevant to studying physics. Similarly, the physics items represent the prior physics knowledge considered essential for studying physics.

To interpret differences between cohorts meaningfully, it is essential that the conditions of test administration have remained as comparable as possible across time. Therefore, all SETs were conducted under standardized conditions: The SETs are carried out in the first two weeks of the first semester in the form of a paper-and-pencil test [7]. In each test administration (in 1978, 2013, and 2023), the test duration was 90 min and consistently included demographics, mathematics items, and physics items. After administration, local staff at the participating universities carried out local scoring based on a scoring manual. The scoring categories are "correct response," "incorrect

TABLE II. Interrater reliability [$M(\kappa)$] between local and central scorings in mathematics and physics across both cohorts (2013/2023).

Scoring	$M(\kappa)$ mathematics	$M(\kappa)$ physics
Local 2013–central 2013	0.91	0.85
Local 2023–central 2023	0.92	0.81
Local 2013–central 2023	0.95	0.93

response," and "omitted response." After the SET booklets were sent back to the authors, a central scoring was conducted on 10% (or at least 10 tests) of each university's test booklets to calculate an interrater reliability and ensure scoring accuracy.

Following the procedure of Wirtz and Casper [51], as employed by Buschhuter *et al.* [7], the interrater reliability was calculated with both scorings assumed to be pseudoraters, because multiple people could have scored the tests at each university. The procedure also accounts for the number of students at each university, making it more appropriate than an unweighted interrater reliability (kappa value, κ).

Three interrater reliabilities were calculated to assess the consistency of scoring within and between the 2013 and 2023 scorings, see Table II.

Given that all interrater reliabilities—within 2013, within 2023, and between 2013 and 2023—fall within the "almost perfect" range [52], this indicates reliable scoring in both years. The stability of the Kappa values over time further suggests a similar level of strictness between 2013 and 2023.

Summarizing, the revised SET 2023 has been shown to validly measure mathematics prior knowledge comparable with SET 2013. It is important to note that SET 2023 (as well as SET 2013 and the first version, SET 1978, see Refs. [7,42,53]) was not designed as a traditional psychometric test. Consequently, conventional benchmarks—such as a mean score of 50%—do not apply. The SET is designed such that all students should be able to solve all items correctly. This approach is primarily because the items were selected by university physics teachers, emphasizing the essential prior knowledge that first-semester students in physics must master to successfully follow lectures and perform well on their initial introductory physics exam.

III. AIM OF THE PRESENT STUDY

The main aim of the present paper is to examine changes in mathematics prior knowledge from 2013 to 2023, building on previous research that highlights its critical role in physics success. By extending the study of Gahrmann *et al.* [21], this paper seeks to explore how mathematics prior knowledge has evolved across both cohorts and to analyze how these changes relate to

corresponding developments in physics prior knowledge. Additionally, by analyzing specific mathematics and physics items, cohort differences are identified, deepening the understanding of trends in the content-related dimension of study preparedness. Additionally, as some universities participated in both 2013 and 2023, it is possible to analyze changes in the average test performance over time within the same institutional contexts to understand to what extent the prior knowledge varies at the same institutions in different years. This enables a control analysis, explaining possible heterogeneity due to structural differences between universities, and thus allows to gain a more detailed understanding of observed general changes.

As such, this paper addresses the following four research questions:

RQ1: To what extent has the mathematics prior knowledge of first-semester physics students in Germany changed between 2013 and 2023?

RQ2: How are the changes in mathematics prior knowledge related to changes in physics prior knowledge over the same period?

RQ3: Within individual institutions, to what extent has mathematics and physics prior knowledge of first-year students changed between the 2013 and 2023 cohorts?

RQ4: Which specific mathematics and physics items show cohort differences, and what do these differences reveal about trends in the content-related dimension of study preparedness?

IV. METHODS

A. Sample

The sample consists of $N = 2251$ students from $N = 27$ German universities in 2013 and $N = 2007$ students from $N = 34$ universities in 2023. An overview of the demographic information is provided in Table III, also given in Ref. [21].

Note that the total number of students decreased from 2013 to 2023. The majority of the students were enrolled as physics majors rather than physics teacher students in both cohorts. Mean school GPA remained stable (2013: $M = 2.0$, $SD = 0.6$; 2023: $M = 1.8$, $SD = 0.6$), and the proportion of students who had attended preparatory

TABLE III. Demographic information of the 2013 and 2023 cohorts. Note that the table is adapted from Ref. [21].

Demographic characteristic	2013	2023
N_{students}	2251	2007
$N_{\text{universities}}$	27	34
Percentage _{Male} Percentage _{Female}	78 21	72 28
School GPA: M (SD)	2.0 (0.6)	1.8 (0.6)
Participation rate in preparatory courses before university	70%	67%

courses prior to university entry was similar as well (2013: 70%; 2023: 67%). The 2013 and 2023 cohorts are comparable in key demographic characteristics. These similarities provide a valid basis for comparing the cohorts' mathematics and physics prior knowledge.

B. Statistical analyses

Before a comparison between the 2013 and 2023 cohorts was performed, the measurement invariance for the $N = 17$ mathematics and $N = 17$ physics items as a subset of the $N = 37$ mathematics and $N = 32$ physics items was again examined using R and the package LAVAAN [46,47]. Here, partial scalar invariance was found for the $N = 17$ mathematics and $N = 17$ physics items, allowing for comparison of the mean scores of both cohorts.

To address the first research question, the sum score for each student was calculated. In line with the procedure applied in 2013, missing responses were treated as incorrect. Given that the test duration was similar and a similarly low number of missing responses was found, this approach was considered acceptable to compare both cohorts, although it may have underestimated students' actual performance in both 2013 and 2023. A Mann-Whitney U test was then conducted to determine whether the distribution of scores differed significantly between the two cohorts. This nonparametric test compares the ranks of the scores and is appropriate when the assumptions of normality for a t test are not met. Given the large sample size, effect size calculations were performed to assess the magnitude of the difference between the distributions.

The effect size was calculated using the following formula:

$$d = \frac{\mu_{2023} - \mu_{2013}}{SD_{\text{pool}}},$$

with

$$SD_{\text{pool}} = \sqrt{\frac{(n_{2013} - 1)s_{2013}^2 + (n_{2023} - 1)s_{2023}^2}{n_{2013} + n_{2023} - 2}}$$

being the pooled standard deviation, μ_i the mean values of the sum scores for the i th cohort, n_i the sample size of the i th cohort, and s_i the standard deviation of the i th cohort $i \in \{2013, 2023\}$. d can be interpreted as a standardized difference between the mean values of both cohorts scaled by the pooled standard deviation. Cohen [54] provided a widely used “rule of thumb” for interpreting effect sizes in (educational) research, which is used to broadly guide the interpretation, that is, an effect size of 0.2 is considered small, 0.5 medium, and 0.8 large. Effect sizes below 0.2 are interpreted as negligible and not of practical interest.

Addressing the second research question, the observed differences in mathematics scores are then compared to

the differences in physics performance reported by Gahrman *et al.* [21].

To investigate the third research question—regarding the stability in student performance at individual universities that participated in both years, 2013 and 2023—we included only universities with a comparable number of participants in both years. This yielded a set of $N = 8$ universities represented both in the 2013 and 2023 data. In each cohort, the sample size per university was over 50. The analyses are thus based on a subsample of $N = 8$ universities, which participated in both years, 2013 and 2023, and had at least 50 students participating in each cohort. For each university, student performance was aggregated into a single institutional mean score, representing the average performance at that institution. These institutional means were then compared across both cohorts using ANOVAs (Analysis of Variance) to examine whether statistically significant differences emerged. Bonferroni-corrected Post-hoc tests (Bonferroni-corrected t tests) were subsequently conducted to identify specific differences between institutions. The Bonferroni correction is applied to control for the increased risk of type I errors (false positives) that occur when multiple comparisons are made.

Regarding the fourth research question, differences in the proportion of correct responses were analyzed for specific mathematics and physics items across the two cohorts. Items exhibiting the largest discrepancies (i.e., differences exceeding 5 percentage points) were examined in terms of their content to identify potential factors contributing to these variations.

V. RESULTS

A. Changes in mathematics prior knowledge (RQ1)

To assess changes in mathematics prior knowledge, we analyzed the scores from the mathematics items administered to the 2013 and 2023 cohorts. Both cohorts' mathematics score distributions are presented in Fig. 1 (top panel). Both distributions cover the full range of sum scores up to the maximum value of 17 and exhibit a similar, left-skewed shape, indicating that most students scored high. A Shapiro-Wilk test confirmed that neither distribution was normally distributed (which was also checked via Q-Q plots). This is consistent with the observed skewness.

Table IV summarizes these statistical characteristics of both mathematics distributions. Although the mean values and standard deviations are reported, these measures should be interpreted with caution because of the skewed nature of the data.

A Mann-Whitney U test revealed a statistically significant difference between the two cohorts [$z(N_1 = 2251, N_2 = 2007) = -3.38, p < 0.001, r = 0.05$]; however, the effect size (Cohen's $d = 0.10, 95\% \text{ CI} = [0.04, 0.16]$) suggests that the practical difference in overall mathematics prior knowledge is negligible [54]. This suggests that,

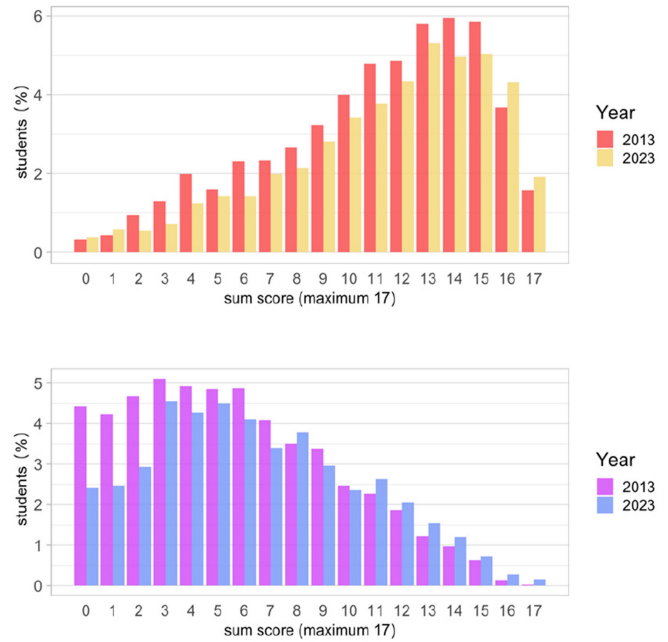


FIG. 1. Top: distribution of students per mathematics sum score for the 2013 cohort (red) and the 2023 cohort (yellow). Bottom: distribution of students per physics sum score for the 2013 cohort (purple) and the 2023 cohort (blue), see [21].

based on sum scores from multiple content areas, mathematics prior knowledge remains stable across cohorts.

B. Changes in mathematics prior knowledge compared to changes in physics prior knowledge (RQ2)

The findings of stable mathematics prior knowledge parallel the findings of Gahrman *et al.* [21], who found small changes in physics prior knowledge (Cohen's $d = 0.21$) within the same cohorts, see Fig. 1 (bottom panel, see also [21]). In mathematics as well as in physics, the median and mean sum scores are similar across time: for mathematics, $Md_{2013} = Md_{2023} = 12$ ($M_{2013} = 10.93, M_{2023} = 11.32$), and for physics, $Md_{2013} = 5$ and $Md_{2023} = 6$ ($M_{2013} = 5.64, M_{2023} = 6.45$).

Mann-Whitney U tests further confirm that the mathematics scores are significantly higher than physics scores in 2013 ($z(N_1 = N_2 = 2322) = 38.02, p < 0.001, r = 0.79$) and 2023 ($z(N_1 = N_2 = 2007) = 33.37, p < 0.001, r = 0.74$).

Comparing the sum score difference between mathematics and physics for 2013 (7 points from 17 overall) and 2023 (6 points from 17 overall), a gap is notable. This gap between mathematics and physics prior knowledge appears to be unchanged over time, as shown in Table IV.

C. Cohort comparison within individual institutions (2013 vs 2023) (RQ3)

To address the third research question—to what extent mathematics prior knowledge and physics prior knowledge changed between the 2013 and 2023 cohorts at the

TABLE IV. Statistics of the 2013 and 2023 mathematics sum score distributions. Note that because of the large sample, normality was also checked via Q-Q plots (plot of quantiles of both distributions against each other). M , mean value; SD, standard deviation; Md , median; MAD, median absolute deviation.

Cohort	N	M (SD)	Md (MAD)	Skewness	Kurtosis	p value Shapiro-Wilk test
2023	2007	11.32 (3.97)	12.00 (4.45)	−0.76	−0.10	<0.001
2013	2251	10.93 (4.00)	12.00 (4.45)	−0.64	−0.41	<0.001

universities participating in both years—we compared cohort performance within individual institutions.

While theoretical frameworks on subject-specific prior knowledge provide valuable insights into students' learning and preparation, they do not explicitly predict temporal changes in prior knowledge across cohorts or account for variability between institutions. Therefore, it is essential to examine changes not only at the aggregated, cross-university level but also within universities that participated in both cohorts.

This focus is important because universities that participated only in one cohort (either 2013 or 2023) may differ in institutional characteristics or student populations, which can influence the overall patterns of change. By comparing only those institutions involved in both years, we reduce potential biases caused by differences in university composition and better isolate temporal changes in prior knowledge. The following analyses are thus based on a subsample of $N = 8$ universities, which participated in both years, 2013 and 2023, and had at least 50 students participating in each cohort.

Overall, we observed only small differences in student performance between the 2013 and 2023 cohorts. The descriptive results are summarized in Fig. 2. On average, the institutional mean scores in mathematics were slightly higher in 2023 ($M = 11.43$, $SD = 1.46$, $Md = 11.65$, $MAD = 1.42$) than in 2013 ($M = 10.95$, $SD = 0.40$, $Md = 10.85$, $MAD = 0.22$). A similar pattern was observed for physics: the mean score across universities in 2023 ($M = 6.34$, $SD = 1.16$, $Md = 6.13$, $MAD = 1.61$) was slightly higher than in 2013 ($M = 5.85$, $SD = 0.42$, $Md = 5.88$, $MAD = 0.54$). An important finding here is the larger heterogeneity of the achieved mathematics and physics scores in 2023, in contrast to 2013, when the universities were put together at each cohort, as shown in Fig. 2.

Furthermore, the one-way ANOVA without repeated measures for mathematics revealed significant main effects of cohort [$F(1, 1876) = 28.70$, $p < 0.001$] and university [$F(7, 1876) = 5.31$, $p < 0.001$]. Importantly, a significant interaction between cohort and university emerged [$F(7, 1876) = 9.59$, $p < 0.001$].

For physics, the one-way ANOVA without repeated measures also showed significant main effects of cohort [$F(1, 1876) = 13.66$, $p < 0.001$] and university [$F(7, 1876) = 3.39$, $p = 0.001$] as well as a significant

cohort by university interaction [$F(7, 1876) = 3.95$, $p < 0.001$].

We estimated eta squared (η^2) as a measure of explained variance for the factors in our ANOVA models. For mathematics, η^2 indicated small effect sizes for year ($\eta^2 = 0.01$), university ($\eta^2 = 0.02$), and the year by university interaction ($\eta^2 = 0.03$). Similarly, for physics, effect sizes were small for year ($\eta^2 = 0.007$), university ($\eta^2 = 0.01$), and the interaction ($\eta^2 = 0.01$). These values suggest that only a negligible proportion of the variance in test scores can be attributed to the cohort, the institution, or their interaction.

These significant interactions indicate that changes over time were not uniform across universities, warranting a closer look at *post hoc* tests within each institution.

The results of the Bonferroni-corrected *post hoc* tests (Bonferroni-corrected t tests) are summarized in Table V and Fig. 3 for mathematics and in Table VI and Fig. 4 for physics. In mathematics, significant differences between the 2013 and 2023 cohorts were found at four out of eight analyzed universities. Specifically, three universities

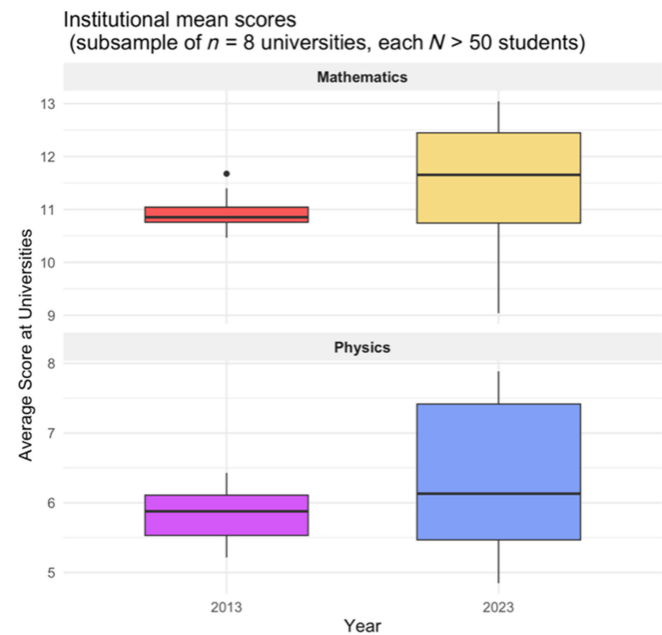


FIG. 2. Boxplots of average student performance (M) in 2013 and 2023, separated (top: Mathematics, bottom: Physics). Only universities with more than 50 participating students and participation in both cohorts were included ($n = 8$ universities).

TABLE V. Table Bonferroni-corrected *post hoc* tests mathematics. d.o.f. denotes the degrees of freedom, *t* the test power of each *post hoc t* test.

University	Estimate	SE	d.o.f.	<i>t</i>	<i>p</i> value	Cohen's <i>d</i>
Uni 1	−0.27	0.72	123	−0.37	1.000	0.06
Uni 2	−2.47	0.43	250	−5.69	<0.001	0.70
Uni 3	1.88	0.63	171	2.97	0.027	−0.45
Uni 4	−1.66	0.43	178	−3.88	0.001	0.46
Uni 5	−2.24	0.34	391	−6.65	<0.001	0.64
Uni 6	1.89	0.79	94	2.39	0.153	−0.45
Uni 7	−0.83	0.40	259	−2.07	0.315	0.24
Uni 8	−0.17	0.67	125	−0.25	1.000	0.04

(Uni 2, Uni 4, and Uni 5) showed significantly higher mean scores in 2023 compared to 2013, while one university (Uni 3) showed significantly lower mean scores in 2023. For the remaining universities (Uni 1, Uni 7, and Uni 8), no statistically significant differences were observed.

In physics, the Bonferroni-corrected *post hoc* tests (Bonferroni-corrected *t* tests) showed statistically significant differences at three out of the eight analyzed universities (Uni 2, Uni 4, and Uni 5), all showing significantly higher mean scores in 2023 compared to 2013. At the other universities, no statistically significant differences were detected. It is important to note that the university numbers are arbitrary to anonymize the universities, but they are matched between mathematics and physics.

These findings suggest that changes in mathematics prior knowledge between cohorts are not uniform but rather

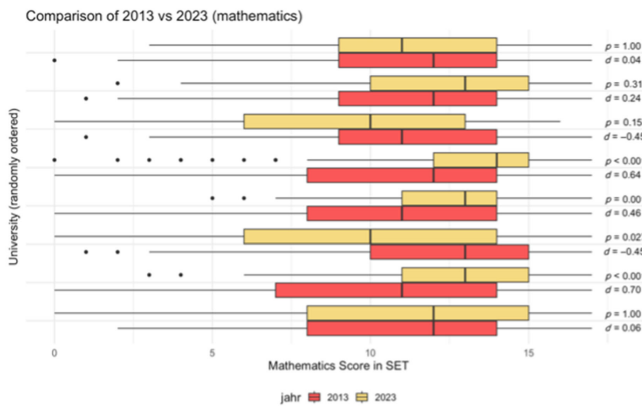


FIG. 3. Comparison of the mathematics performance of students in the 2013 (red) and 2023 (yellow) cohorts at eight randomly arranged universities. The boxplot shows the distribution of mathematics scores in the Study Entrance Test (SET) for each university and cohort. Next to each university, the corresponding *p* values and effect sizes (Cohen's *d*) from the *post hoc* tests are provided, assessing the differences between the cohorts. The *p* values indicate that the differences are statistically significant at several universities, with mathematics performance partly improving (higher mean scores in 2023) or declining. Effect sizes range from “negligible” ($d = 0.04$) to “large” ($d = 0.70$), reflecting varying degrees of change.

TABLE VI. Table Bonferroni-corrected *post hoc* tests physics. Note that d.o.f. denotes the degrees of freedom, *t* the test power of each *post hoc t* test.

University	Estimate	SE	d.o.f.	<i>t</i>	<i>p</i> value	Cohen's <i>d</i>
Uni 1	0.06	0.69	146	−0.09	1.000	−0.01
Uni 2	−2.07	0.56	250	3.72	0.002	0.46
Uni 3	1.26	0.69	150	−1.83	0.551	−0.28
Uni 4	−1.79	0.54	166	3.33	0.009	0.40
Uni 5	−1.71	0.41	414	4.14	<0.001	0.40
Uni 6	0.37	0.74	114	−0.50	1.000	−0.09
Uni 7	0.03	0.51	282	−0.05	1.000	−0.01
Uni 8	−0.09	0.81	105	0.11	1.000	0.02

dependent on institutional contexts: While students at some universities showed improved mathematics scores, other scores decreased. In physics, a more uniform trend can be seen: Some universities showed improved physics scores, whereas at other universities, no change is visible.

D. Different scores in mathematics and physics items (RQ4)

To gain deeper insight into cohort differences, we examined item-level performance across mathematics and physics. Figure 5 displays scatterplots of the proportion of correct responses in 2013 vs 2023—mathematics in the top plot, physics in the bottom. A diagonal reference line with a slope of 1 represents identical performance in both cohorts. Items positioned close to this line were answered similarly by both cohorts, while items above the line were

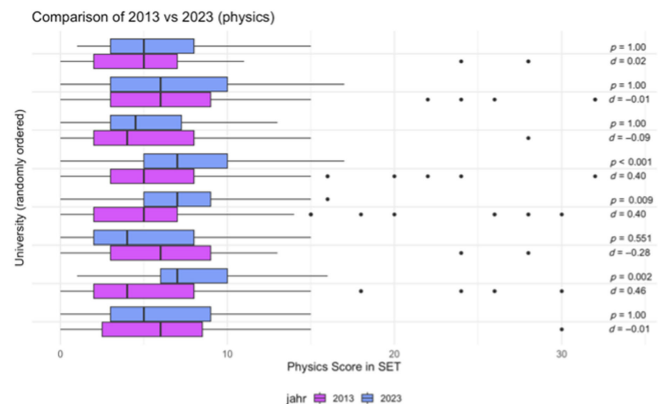


FIG. 4. Comparison of the physics performance of students in the 2013 (purple) and 2023 (blue) cohorts at eight randomly arranged universities. The boxplot shows the distribution of physics scores in the Study Entrance Test (SET) for each university and cohort. Next to each university, the corresponding *p* values and effect sizes (Cohen's *d*) from the *post hoc* tests are provided, assessing the differences between the cohorts. The *p* values indicate that the differences are statistically significant at several universities, with physics performance improving (higher mean scores in 2023). Effect sizes range from “negligible” ($d = 0.01$) to “moderate” ($d = 0.46$), reflecting improvements.

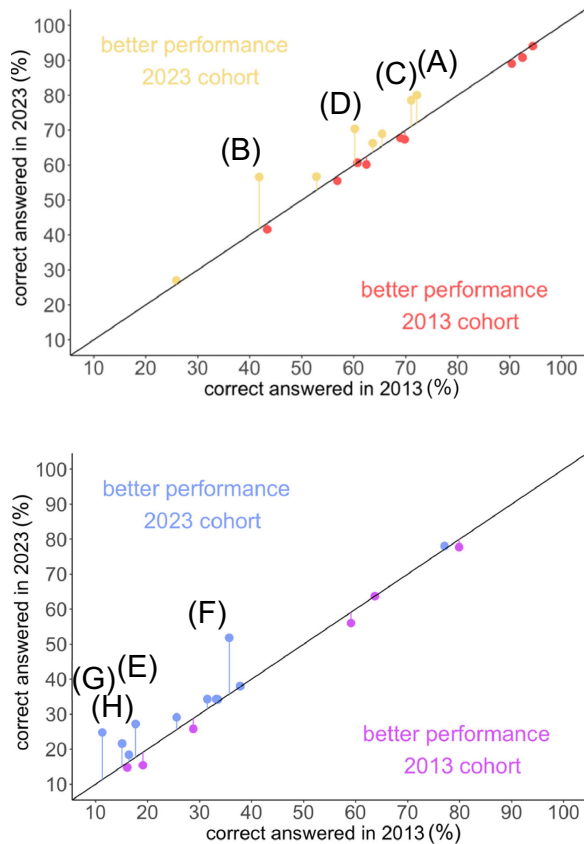


FIG. 5. Scatterplots comparing item performance between the 2013 and 2023 cohorts. Each point is an item; the diagonal indicates equal performance. Colors show which cohort performed better: yellow = 2023 (math), red = 2013 (math), blue = 2023 (physics), pink = 2013 (physics); see Fig. 1. Top: Math anchor items. Bottom: Physics anchor items. Letters refer to items in Fig. 6.

solved more accurately in 2023, suggesting improved prior knowledge. Items below the line were answered more accurately in 2013, indicating a relative decline in those items. In general, most items in both domains cluster around the line, indicating stable overall performance between cohorts. Although some items deviate from this line—mostly in favor of the 2023 cohort—the magnitude of these differences is relatively small. The largest observed difference was 14.8 percentage points in mathematics (average: 2.3%) and 16.1 percentage points in physics (average: 2.5%). These modest differences suggest overall stability in students’ preparedness, with only select items showing meaningful variation.

For a more detailed analysis, we identified items with a performance difference greater than 5 percentage points. This threshold resulted in eight items—four from mathematics and four from physics—depicted in Fig. 6 along with their respective correct-answer rates and differences (Δ). In mathematics, the four items with notable cohort differences (labeled A to D in Figs. 5 and 6) span three key

content areas: vector calculus, functions, and integration. Item A focused on vector calculus and required knowledge of vector algebra, particularly the conditions for perpendicularity. Students in the 2023 cohort performed better on this item, suggesting a modest improvement in conceptual understanding of vector representations. Items B and C both addressed the domain of functions. In item B, students were asked to draw a linear function given 2 points and determine the corresponding equation. The accuracy among students in 2023 points to a slightly enhanced procedural competence in constructing linear functions and the usage of equations, which may reflect a stronger curricular emphasis on such skills in secondary education. However, despite this improvement, it is noteworthy that fewer than 60% of incoming students were able to solve the item correctly, indicating that challenges with representational translation remain substantial. In contrast, item C involved translating a symbolic equation into a graphical (iconic) representation. Here, performance also increased in the 2023 cohort, suggesting gains in the ability to shift between different representational formats. Item D addressed basic integration and required students to apply elementary calculus rules. The better performance of the 2023 cohort suggests an improvement in foundational procedural knowledge in this area.

Turning to physics, the four items with the most substantial performance differences (items E to H, again shown in Figs. 5 and 6) covered mechanics, electronics and electrodynamics, and wave phenomena. Item E, drawn from mechanics, involved differentiating a function with respect to time and understanding the concept of velocity. Students in 2023 showed higher levels of accuracy, indicating improved procedural fluency in applying mathematical operations to physical quantities, although this improvement may not necessarily reflect deeper conceptual understanding. Items F and G pertained to electricity and electrodynamics. In Item F, students were asked to sketch a simple electric circuit using components such as resistors. The stronger performance of the 2023 cohort may reflect greater familiarity with circuit diagrams, possibly due to increased practical exposure in school physics lessons. Item G focused on conceptual knowledge in electromagnetism. While performance was still relatively low, the 2023 cohort demonstrated a slight increase in correctness, pointing to modest gains in conceptual understanding, even in this traditionally challenging domain. A similar trend was observed in item H, which dealt with wave phenomena. This item required students to understand and graphically represent wave interference—a conceptually demanding task. Compared to 2013, the 2023 cohort showed improved performance, albeit on a low level, suggesting incremental progress.

VI. DISCUSSION

Motivated by persistently high dropout rates in physics and the well-documented role of insufficient prior

When is the scalar product of two vectors equal to zero?	2013: 72.1 % 2023: 80.0 % $\Delta = 7.9 \%$	(a)	The trajectory of a mass point is described by the path curve $\vec{r}(t) = 3\vec{g}t^2 - \vec{v}_0t + \vec{r}_0$. What is the velocity of this mass point at time t ?	2013: 17.7 % 2023: 27.2 % $\Delta = 9.5 \%$	(e)
(a) Draw the line passing through the points $P_1 = (0,1)$ and $P_2 = (-2,-1)$ in the coordinate system. (b) Determine the equation of this line in terms of the coordinates u and v .	2013: 41.8 % 2023: 56.6 % $\Delta = 14.8 \%$	(b)	Use a circuit diagram to illustrate how to measure electric current with a measuring device. Include a resistor, a voltage source, and the measuring device in your diagram.	2013: 35.7 % 2023: 51.8 % $\Delta = 16.1 \%$	(f)
Draw a sketch of the graph of the following function, if you are familiar with it. (Please plot it in the given coordinate systems.) Example: $y = x^2$	2013: 71.1 % 2023: 78.6 % $\Delta = 7.5 \%$	(c)	Briefly describe the principle of an induction process.	2013: 11.3 % 2023: 24.8 % $\Delta = 13.5 \%$	(g)
$\int_{-1}^{+1} x^3 dx =$	2013: 60.2 % 2023: 70.4 % $\Delta = 10.2 \%$	(d)	Explain the interference of light in a double-slit experiment using Huygens' principle. Provide a sketch to illustrate the explanation.	2013: 15.1 % 2023: 21.6 % $\Delta = 6.5 \%$	(h)

FIG. 6. Anchor items in mathematics (left, A–D) and physics (right, E–H) with a difference in both cohorts larger than 5% in correct answers. With each item, the rate of correct answers in each cohort, as well as their difference (denoted Δ).

knowledge in mathematics and physics as a principal cause of attrition in physics degree programs, we undertook an analysis of how incoming students' prior knowledge has evolved between 2013 and 2023.

A. RQ1 and RQ2: Changes in prior knowledge in mathematics and physics (2013–2023)

This study investigated whether the mathematics and physics prior knowledge of first-semester physics students in Germany has changed between 2013 and 2023, and how these changes relate to each other. We observed improvements in mathematics and physics [21] on a general level when all participating universities in 2013 and 2023 were included in the analysis. Students in 2023 outperformed the 2013 cohort slightly in both mathematics (Cohen's $d = 0.10$) and physics (Cohen's $d = 0.21$).

Interpreted through the framework of Yeager *et al.*, these effect sizes correspond to roughly half a year (for the mathematics changes) and a full year (for the physics changes) of classroom learning at the ninth-grade level [55]. Although these effect sizes are modest and would traditionally be labeled as “small” [54], they might suggest that students entering physics programs in 2023 are somewhat better prepared than their 2013 counterparts.

The better performance in the 2023 cohort differs from the interpretable results of Buschhüter *et al.*, who found a significant decline in physics and mathematics prior knowledge between 1978 and 2013 (Cohen's $d = -0.53$ in physics; $d = -0.20$ in mathematics) [7,15]. Our findings may therefore indicate a stabilization or slight recovery since 2013. However, caution is required in comparing across decades and especially both studies: while the 2013–2023 comparison is based on 17 anchor items in mathematics and 17 anchor items in physics, the 1978–2013 comparison used larger item sets (43 items in mathematics and 38 items in physics).

The improved performance of the 2023 cohort compared to 2013 is striking, given the overall decline in school achievement documented in large-scale assessments like PISA and NAEP [20,35]). One possible explanation lies in the specific student population: first-semester physics students represent a selected, likely more academically resilient group whose competencies may have remained stable—or even improved—despite broader systemic challenges.

Supporting this interpretation, results from Büchele and Feudel reported stable or slightly improved mathematical competencies among economics students over the past decade [56]. Like in our study, they observed gains in areas such as mathematization and the use of representations, while persistent difficulties with symbolic and procedural tasks remained—a pattern mirrored in our item-level analyses (see Sec. VIC).

Nevertheless, absolute achievement levels remain low. Students in 2023 only solved 67% of math items and 38% of physics items on average, despite the fact that all test items were designed to reflect essential competencies as defined by the German National Educational Standards [57,58]. This substantial gap between target (all items correct as the items address essential prior knowledge) and actual performance indicates foundational weaknesses that remain critical for university success.

B. RQ3: Cohort comparison within individual institutions

To examine whether the changes in mathematics and physics prior knowledge between the 2013 and 2023 cohorts were consistent across institutions, we analyzed data from the eight universities that participated in both years of measurement. This analysis complements the findings on overall cohort differences (RQ1), where we observed no changes in average mathematics and physics scores over time.

The descriptive results revealed a broader range of student performance in 2023 compared to 2013. This increased heterogeneity suggests that, even though average scores remained stable, the variability of prior knowledge among incoming students has grown. Consequently, instructors may face more heterogeneous cohorts, which could require more differentiated teaching strategies and support systems.

Addressing RQ3 in more detail, we conducted ANOVAs, one for mathematics and one for physics prior knowledge, with interaction terms to assess whether the observed changes differed between individual universities. This step was crucial because the aggregated analysis (RQ1) included universities that participated in only one of the two measurement years, potentially masking institution-specific trends.

The ANOVA results showed significant interactions between cohort and university for both mathematics and physics, indicating that the changes over time were not uniform across institutions. Although the overall effect sizes were small, *post hoc* tests revealed meaningful patterns at the local level.

In mathematics, prior knowledge changes varied across universities, with some showing significant increases, others showing decreases, and several showing no change. In physics, improvements were more consistent but limited to a few universities, indicating that gains were driven by specific institutions rather than a general trend.

Taken together, these findings nuance the results from RQ1. While aggregated analyses suggested no overall differences between the 2013 and 2023 cohorts in mathematics and physics prior knowledge, the institutional analyses for RQ3 revealed considerable variation across universities. In mathematics, local increases and decreases in performance appear to have canceled each other out at the overall level. For physics, however, the slight overall increase may reflect actual improvements at universities.

Summarizing RQ1–Q3, mathematics and physics prior knowledge remained broadly stable across cohorts when analyzed at the national level but varied notably across institutions. These results emphasize the importance of considering both aggregated trends and local institutional developments when examining changes in student preparedness over time.

C. RQ4: Item-level differences and content-specific trends

To gain deeper insight into the changes in prior knowledge, we conducted an item-level analysis across both cohorts. Notably, eight test items (four in mathematics and four in physics) demonstrated a performance improvement exceeding 5%, serving as important examples of areas where student preparedness has meaningfully increased from 2013 to 2023. Meanwhile, most items (13 in mathematics and 13 in physics) showed little

to no change, indicating stable performance across cohorts. It is important to note that the test primarily targeted declarative and basic procedural knowledge; more complex problem-solving tasks were not represented in the items used for the present study.

In mathematics, performance improvements were observed in items requiring declarative skills, such as drawing linear functions or applying basic integration rules. These developments may reflect curricular shifts toward an increased focus on fluency in core mathematical operations. For example, improved outcomes on vector calculus items (e.g., Item A, Fig. 6) may point to greater emphasis on spatial representations in the upper secondary curriculum. However, whether this reflects an actual curricular shift toward spatial representations would require normative evidence from curriculum analyses or textbook studies.

In physics, a comparable pattern emerged. Students performed better on items requiring declarative knowledge, such as differentiating functions (Item E, Fig. 6) or assembling basic electric circuits (Item F, Fig. 6). However, items with a focus on more conceptually demanding knowledge—such as understanding electromagnetic phenomena (Item G, Fig. 6) or wave interference (Item H, Fig. 6)—showed smaller gains. This is in line with previous findings that more conceptual understanding in physics tends to develop more slowly and is harder to promote [59].

Taken together, these item-level trends suggest that while declarative and foundational procedural aspects of prior knowledge have improved across cohorts, challenges remain in fostering more integrated conceptual understanding. The absence of complex problem-solving items in the test underscores that even within a relatively narrow cognitive domain, meaningful shifts in performance can be observed over time.

VII. LIMITATIONS

Before drawing implications, it is important to consider limitations that may influence the generalizability of our results. In this paper, prior knowledge in mathematics and physics was compared between two cohorts, but it is important to note that no causal conclusions can be drawn from the data. The observed differences may reflect various contributing factors, such as those described in the discussion, but without further investigation, the causal mechanisms behind these changes remain unclear.

The lack of general changes across cohorts does not necessarily imply that no changes have occurred. Even small effect sizes can have significant educational implications, and the absence of large-scale changes does not preclude the possibility that other factors may have influenced individual competencies in specific areas. The discussion of effect sizes should be considered in this context.

Some of our analyses (on RQ3) were based on a subsample of universities that participated in both 2013 and 2023, which allowed for a more controlled comparison over time.

A key limitation of RQ1 is the heterogeneity in university participation across cohorts, which complicates the interpretation of cohort-related changes in subject-specific prior knowledge. Therefore, we performed analyses on a subsample of universities participating in both 2013 and 2023 to enable controlled comparisons (RQ3). The results revealed significant interactions between cohort and institution. This indicates that changes over time were not uniform but varied considerably between universities.

Therefore, the overall negligible effect sizes observed at the aggregated level may mask substantial local increases or decreases in mathematics and physics prior knowledge of students at individual institutions. This variation limits the generalizability of cohort comparisons based solely on pooled data, as such analyses may overlook important institution-specific trends.

From a practical perspective, general trends in prior knowledge across cohorts (interpreted as negligible changes) may be most relevant for policymakers and educational stakeholders interested in broad systemic changes. In contrast, the institution-specific variations (partly large effect sizes) revealed by our analyses are particularly important for university administrators and faculty, who may need to tailor interventions and support measures to their local context.

Additionally, the scope of our study is inherently limited by the nature of the assessment instruments used. Specifically, the analysis is based on 17 anchor items in mathematics and 17 anchor items in physics, which may not fully capture the entire spectrum of competencies required for first-year physics students. As discussed in the project background, this selection reflects a valid but necessarily narrow subset of relevant content. Accordingly, conclusions drawn from item-level analyses should be interpreted with caution and understood as indicative rather than exhaustive. Furthermore, one limitation pertains to the range of content areas assessed with the anchor items: while the anchored items of SET 2013 and SET 2023 effectively evaluate core knowledge domains typically emphasized at the beginning of classical physics curricula, they do not include more advanced or interdisciplinary topics such as computational skills, data literacy, or broader competencies that have gained importance through the integration of technology and innovative teaching methods. As a result, the findings are best understood as reflecting developments in foundational knowledge and shifts in average performance, while generalizability to modernized or interdisciplinary programs remains limited. A future readministration of the SET 2023, incorporating newly developed items targeting competencies such as computational skills and data literacy, as well as extended

mathematics knowledge, may help to overcome this limitation by enabling comparisons beyond foundational knowledge domains.

Although the assessments were developed by subject-matter experts and provide meaningful resolution for detecting cohort-level differences, several limitations of the metrics should be noted. The mathematics assessment includes items that many students answer correctly, which may produce ceiling effects and reduce discrimination among higher-performing students. Conversely, the physics assessment shows generally low performance, which may limit sensitivity to differences at the lower end of the scale. In addition, the limited number of items restricts the coverage of some content domains, and the assessments cannot capture all aspects of, e.g., problem-solving strategies, reasoning processes, or interdisciplinary knowledge that may be relevant for first-year physics students.

It is important to note that, as students progress through their first semester in physics, other factors—such as perceived stress or sense of belonging—become increasingly more important for their persistence. These factors relate more to subjective aspects of study success, see e.g., Ref. [14]. As students' preparation—and within this, the prior knowledge—plays a highly important role for success in a university physics study program, in this paper, we only focused on the mathematics and physics prior knowledge.

VIII. IMPLICATIONS

Based on the findings of this study, several implications can be derived for research and educational practice. First, it is surprising that no significant changes were observed in students' prior knowledge between 2013 and 2023. It is surprising because other large-scale studies, such as NAEP and PISA, indicate changes and declines in knowledge [19,60]. Given these declines, one might have anticipated a measurable decline in foundational knowledge.

In interpreting the observed cohort differences in mathematics and physics prior knowledge, we reported conventional effect sizes (Cohen's d). According to Cohen's widely used benchmarks— $d = 0.2$ (small), $d = 0.5$ (medium), and $d = 0.8$ (large)—most differences in our study would be seen as “small” or “negligible.” Cohen himself [54], however, pointed out that these thresholds are only rough guidelines and should be adjusted depending on the research context. Small effects may still have practical relevance, especially in fields like education, where cumulative or systemic impacts are possible.

However, both models have limitations when applied to our context and raise questions about what differences we found in our analysis. Cohen's benchmarks are based on more controlled psychological experiments [54], while Yeager's interpretations relate to intervention studies aiming for systemic change [55]. Our study focuses on naturally occurring differences in student cohorts' prior

knowledge—a situation that neither model fully captures. This underlines the need for a critical reflection on the adequacy of standardized thresholds for educational diagnostics and entrance assessments. We recommend interpreting the differences in our study found between all universities participating in 2013 and 2023 as small to no differences between the two cohorts.

The considerations regarding the adequacy of Cohen's or Yeager's interpretation of effect sizes in this context suggest that future research should move beyond simply applying generalized effect size benchmarks. A more meaningful approach would be to examine what a certain difference in prior knowledge implies for students' academic success. For example, how does an improvement of 0.20 standard deviation in mathematics or physics prior knowledge affect the probability of passing key exams in the first semester?

Answering such questions requires longitudinal studies that explicitly link prior knowledge with academic outcomes. By establishing empirical relationships between effect sizes and meaningful educational benchmarks—such as study success, retention rates, or exam performance—future research could define more adequate and context-sensitive thresholds.

In this sense, both Cohen's conservative view and Yeager's more liberal interpretation provide valuable points of reference but should not replace critical, context-driven analyses. We advocate for effect size interpretations that are directly tied to concrete educational outcomes rather than abstract statistical conventions. However, the stability observed in our findings suggests that, at least in the domains assessed, students' preparedness remained relatively unaffected. A possible explanation is that the cohort entering university in 2023 had already acquired most of the relevant knowledge—specifically, the content targeted by the 17 mathematics items and 17 physics items—prior to the onset of the pandemic and the first lockdown. This would imply that the competencies captured by the test largely reflect content typically taught before the final years of secondary education and were thus less vulnerable to pandemic-related educational losses. Nonetheless, this interpretation should be treated with caution, as it relies on assumptions about the timing of content acquisition and does not account for potential qualitative shifts in how knowledge was acquired or retained under pandemic conditions. This counterintuitive result warrants further investigation. While our study focused on foundational mathematics and physics knowledge, it is conceivable that other domains—such as computational skills or data literacy—were more substantially affected by the pandemic. Moreover, the apparent stability in subject-specific prior knowledge may not generalize across all student populations. Existing research suggests that students from socio-economically disadvantaged backgrounds experienced disproportionately greater learning losses during

the pandemic [61]. However, it is also possible that such students are underrepresented among those who eventually enroll in physics degree programs—raising the question of whether our sample captures a more resilient, potentially more privileged subset of the student population. Moreover, students who choose to pursue a physics degree are typically those with a genuine interest in the subject and, in most cases, at least moderate prior achievement in mathematics and physics. This selection effect could partially account for the observed stability and underscores the need for future studies that explicitly examine heterogeneity within incoming student cohorts.

Furthermore, our findings underscore the importance of ongoing assessment of students' prior knowledge. Regularly updating and analyzing such data can inform curriculum development and enable timely support mechanisms, ensuring that educational institutions are better prepared to address future disruptions to learning. Regarding future usages of SETs, it is important to note that the 17 mathematics and 17 physics items used in this study were selected to ensure comparability across cohorts. However, as outlined in Sec. VII, the metrics have some limitations. While the SET was designed so that all students should be able to solve all items, future studies could consider supplementing with additional items to better cover the full range of students' prior knowledge. In mathematics, this could help capture differences among higher-performing students, and in physics, items could help capture differences among lower-performing students. Moreover, for detailed analyses on particular constructs (such as problem solving or comprehensive knowledge on mechanics) instruments would need to be extended by items covering these constructs more broadly. Hereby, it is important that the original items are retained as anchors to allow comparisons with the results presented here.

Moreover, our findings of constant prior knowledge, even though other studies indicate a decline in general knowledge, raise questions about long-term trends. Future research could explore whether a decline in prior knowledge will become evident in the next decade, with international assessments like PISA potentially revealing future shifts in student preparedness [35,60]. It is essential for future research to explore whether the observed trends are a result of broader educational reforms, changes in instructional strategies, or other external influences, providing a clearer understanding of the forces shaping students' knowledge and preparedness. It is therefore reasonable to continue the study in the sense of trend studies such as PISA to explore changes and fluctuations in the study-relevant prior knowledge for physics studies. An additional focus can be placed on specific changes in school curricula, such as focus shifts in the last school years or specific regulations in the Abitur (A-level examination).

ACKNOWLEDGMENTS

We sincerely thank all the students who took part in the study. Our gratitude also goes to the university faculty members whose openness and cooperation were essential for the successful completion of this research. Many thanks to William “Bill” Boone for proofreading the manuscript and improving its readability.

D. G.: conceptualization, methodology, software, formal analysis, resources, data curation; writing—original draft;

writing—review and editing; visualization; I. N.: supervision, writing—review and editing; A. B.: supervision, writing—review and editing.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

-
- [1] A. M. Porter, R. Y. Chu, and R. Ivie, *Attrition and Persistence in Undergraduate Physics Programs* (American Institute of Physics, College Park, MD, 2024).
 - [2] U. Heublein, C. Hutzsch, and R. Schmelzer, *Die Entwicklung der Studienabbruchquoten in Deutschland*, DZHW Brief (DZHW, Hannover, 2022).
 - [3] National Center for Education Statistics, Undergraduate Retention and Graduation Rates. Condition of Education. U.S. Department of Education, Institute of Education Sciences (2022), Retrieved from <https://nces.ed.gov/programs/coe/indicator/ctr>.
 - [4] A. Delahay, M. Lovett, D. Anderson, and S. Sen, Components of the preparation gap for physics learning vary in two learner groups, *Phys. Rev. Phys. Educ. Res.* **19**, 020122 (2023).
 - [5] S. Sorge, S. Petersen, and K. Neumann, Die Bedeutung der Studierfähigkeit für den Studienerfolg im 1. Semester in Physik, *Z. Didakt. Naturwiss.* **22**, 165 (2016).
 - [6] E. W. Burkholder, G. Murillo-Gonzalez, and C. Wieman, Importance of math prerequisites for performance in introductory physics, *Phys. Rev. Phys. Educ. Res.* **17**, 010108 (2021).
 - [7] D. Buschhüter, C. Spoden, and A. Borowski, Physics knowledge of first semester physics students in Germany: A comparison of 1978 and 2013 cohorts, *Int. J. Sci. Educ.* **39**, 1109 (2017).
 - [8] J. Stewart, J. Hansen, and E. Burkholder, Visualizing and predicting the path to an undergraduate physics degree at two different institutions, *Phys. Rev. Phys. Educ. Res.* **18**, 020117 (2022).
 - [9] F. Thiel, D.-P. I. Blüthmann, D.-S. M. Richter, and D.-P. N. Csonka, *Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen* (Freie Universität Berlin, Berlin, 2006).
 - [10] G. Kuh, J. Kinzie, J. Buckley, B. Bridges, and J. Hayek, What matters to student success: A review of the literature, Report No. NPEC 2006-1, National Postsecondary Education Cooperative, 2006.
 - [11] C. Konegen-Grenier and C. Kuhlmann, *Studierfähigkeit Und Hochschulzugang* (Deutscher Instituts-Verlag, Köln, 2002).
 - [12] S. Salehi, E. Burkholder, G. P. Lepage, S. Pollock, and C. Wieman, Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics, *Phys. Rev. Phys. Educ. Res.* **15**, 020114 (2019).
 - [13] M. Solis-Foronda, Predictors of students’ knowledge in general physics, *Univ. J. Educ. Res.* **8**, 3630 (2020).
 - [14] J. Fleischer, D. Leutner, M. Brand, H. Fischer, M. Lang, P. Schmiemann, and E. Sumfleth, Vorhersage des Studienabbruchs in naturwissenschaftlich-technischen Studiengängen, *Z. Erzieh.* **22**, 1077 (2019).
 - [15] D. Buschhüter, C. Spoden, and A. Borowski, Mathematische Kenntnisse und Fähigkeiten von Physikstudierenden zu Studienbeginn, *Z. Didakt. Naturwiss.* **22**, 61 (2016).
 - [16] E. W. Burkholder and C. E. Wieman, What do AP physics courses teach and the AP physics exam measure?, *Phys. Rev. Phys. Educ. Res.* **15**, 020117 (2019).
 - [17] A. Gerdes, S. Halverscheid, and S. Schneider, Teilnahme an mathematischen Vorkursen und langfristiger Studienerfolg. Eine empirische Untersuchung, *J. Math. Didakt.* **43**, 377 (2022).
 - [18] R. Henderson, C. Zabriskie, and J. Stewart, Rural and first generation performance differences on the force and motion conceptual evaluation, presented at PER Conf. 2018, Washington, DC, [10.1119/perc.2018.pr.Henderson](https://doi.org/10.1119/perc.2018.pr.Henderson).
 - [19] NAEP mathematics: National trends and student skills, https://www.nationsreportcard.gov/reports/mathematics/2024/g4_8/national-trends/?grade=8.
 - [20] *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*, edited by D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller, and K. Reiss (Waxmann Verlag GmbH, Münster, Germany, 2023).
 - [21] D. Gahrman, I. Neumann, and A. Borowski, Comparing prior knowledge of first-semester physics students between the cohorts of 2013 and 2023, in *Book of extended abstracts of the 4th World Conference on Physics Education. Embracing Changes Together*, edited by M. Kaczmarek and D. Sokołowska (Astronomy and Applied Computer Science Jagiellonian University, Kraków, Poland, 2024), pp. 254–255, ISBN 978-83-958051-6-5.

- [22] E. W. Burkholder, K. Wang, and C. Wieman, Validated diagnostic test for introductory physics course placement, *Phys. Rev. Phys. Educ. Res.* **17**, 010127 (2021).
- [23] S. Ufer, Studierfähigkeit als eine Zieldimension von Mathematikunterricht in der gymnasialen Oberstufe. Konzepte, Modelle und Beitrag des Mathematikunterrichts, in *Das Fach Mathematik in der Gymnasialen Oberstufe*, edited by T. Rolfes, S. Rach, S. Ufer, and A. Heinze (Waxmann Verlag GmbH, Münster; New York, 2022), pp. 75–102.
- [24] D. Hewagallage, E. Christman, and J. Stewart, Examining the relation of high school preparation and college achievement to conceptual understanding, *Phys. Rev. Phys. Educ. Res.* **18**, 010149 (2022).
- [25] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009).
- [26] S. Z. Lahme, J. O. Cirkel, L. Hahn, J. Hofmann, J. Neuhaus, S. Schneider, and P. Klein, Enrollment to exams: Perceived stress dynamics among first-year physics students, *Phys. Rev. Phys. Educ. Res.* **20**, 020127 (2024).
- [27] T. Binder, A. Sandmann, B. Sures, G. Friege, H. Theysen, and P. Schmiemann, Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression, *Int. J. STEM Educ.* **6**, 33 (2019).
- [28] M. S. Feser, I. Haak, and T. Rabe, Sense of belonging among first-year physics students in Germany: Exploring intergroup differences and correlations, *Eurasia J. Math. Sci. Technol. Educ.* **19**, em2345 (2023).
- [29] J. P. Zwolak, R. Dou, E. A. Williams, and E. Brewé, Students' network integration as a predictor of persistence in introductory physics courses, *Phys. Rev. Phys. Educ. Res.* **13**, 010113 (2017).
- [30] M. Franklin and E. Brewé, What correlates with persistence of women in physics?, *Phys. Rev. Phys. Educ. Res.* **21**, 010115 (2025).
- [31] D. Thomas and J. S. Brown, *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change* (CreateSpace, Lexington, KY, 2011), Vol. 219.
- [32] F. Almasri, Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research, *Res. Sci. Educ.* **54**, 977 (2024).
- [33] G. Roy, S. Sikder, and L. Danaia, Adopting scientific literacy in early years from empirical studies on formal education: A systematic review of the literature, *Int. J. STEM Educ.* **12**, 26 (2025).
- [34] U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 2009–2024 Science Assessments.
- [35] OECD, *PISA 2022 Results (Volume I): The State of Learning and Equity in Education* (OECD, Paris, 2023).
- [36] OECD, *PISA 2022 Technical Report* (OECD, Paris, 2024).
- [37] OECD, *Do Adults Have the Skills They Need to Thrive in a Changing World?: Survey of Adult Skills 2023* (OECD, Paris, 2024).
- [38] A. E. Lawson, The development and validation of a classroom test of formal reasoning, *J. Res. Sci. Teach.* **15**, 11 (1978).
- [39] L. Bao, Y. Xiao, K. Koenig, and J. Han, Validity evaluation of the Lawson Classroom Test of scientific reasoning, *Phys. Rev. Phys. Educ. Res.* **14**, 020106 (2018).
- [40] R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [41] G. Zavala, S. Tejeda, P. Barniol, and R. J. Beichner, Modifying the test of understanding graphs in kinematics, *Phys. Rev. Phys. Educ. Res.* **13**, 020111 (2017).
- [42] F. Krause and A. Reiners-Logothetidou, Der bundesweite Studieneingangstest Physik 1978 (1981).
- [43] D. Buschhüter and A. Borowski, Modellierung von Eingangsanforderungen für Das Studienfach Physik, in *Naturwissenschaftliche Bildung Zwischen Science- und Fachunterricht*, edited by S. Bernholt (Gesellschaft für Didaktik der Chemie und Physik (GDGP), München, 2014), pp. 540–542.
- [44] D. Gahrman, I. Neumann, and A. Borowski, Welche Mathematik- und Physikaufgaben sollten Studierende lösen können, um erfolgreich ins Physikstudium zu starten?—Die Perspektive von Physikdozierenden, *Phydid A* (to be published).
- [45] C. Deeken, I. Neumann, and A. Heinze, Mathematical prerequisites for STEM programs: What do university instructors expect from new STEM undergraduates?, *Int. J. Res. Undergrad. Math. Educ.* **6**, 23 (2020).
- [46] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2024).
- [47] Y. Rosseel, LAVAAN: An R package for structural equation modeling, *J. Stat. Softw.* **48**, 1 (2012).
- [48] F. Hälsig, *Branchenübergreifende Analyse des Aufbaus einer starken Retail Brand* (Gabler, Wiesbaden, 2008).
- [49] B. M. Byrne, Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument, *Int. J. Nondestruct. Test.* **1**, 55 (2001).
- [50] D. L. Streiner, Starting at the beginning: An introduction to coefficient alpha and internal consistency, *J. Pers. Assess.* **80**, 99 (2003).
- [51] M. A. Wirtz and F. Caspar, *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* (Hogrefe, Verlag für Psychologie, Göttingen Bern Toronto Seattle, 2022).
- [52] J. R. Landis and G. G. Koch, The measurement of observer agreement for categorical data, *Biometrics* **33**, 159 (1977).
- [53] F. Krause and A. Reiners-Logothetidou, Folgerungen aus dem Studieneingangstest Physik 1978: Vorschläge zum Abbau der festgestellten Defizite, *Phys. J.* **37**, 295 (1981).
- [54] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, New York, 1988).
- [55] D. S. Yeager et al., A national experiment reveals where a growth mindset improves achievement, *Nature (London)* **573**, 364 (2019).

- [56] S. Büchele and F. Feudel, Changes in students' mathematical competencies at the beginning of higher education within the last decade at a German University, *Int. J. Sci. Math. Educ.* **21**, 2325 (2023).
- [57] Kultusministerkonferenz (Hrsg.), Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (2012), https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf.
- [58] Kultusministerkonferenz (Hrsg.), Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Physik für die Allgemeine Hochschulreife (2020), https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Physik.pdf.
- [59] B. R. Wilcox, S. J. Pollock, and D. R. Bolton, Retention of conceptual learning after an interactive introductory mechanics course, *Phys. Rev. Phys. Educ. Res.* **16**, 010140 (2020).
- [60] OECD, *PISA 2022 Results (Volume II): Learning During—and From—Disruption* (OECD, Paris, 2023).
- [61] European Commission. Directorate General for Education, Youth, Sport and Culture, *Covid-19 Learning Deficits in Europe: Analysis and Practical Recommendations: Analytical Report* (Publications Office, Luxembourg, 2023).