

Comparing conceptual learning in Zoom and in-class sections of college physics

Ramesh Dhungana^{✉*}

Department of Physics, *University of Colorado Denver*,
Denver, Colorado 80217, USA



(Received 15 May 2025; accepted 24 October 2025; published 12 November 2025)

The rapid shift to remote instruction during the COVID-19 pandemic required widespread adoption of synchronous virtual modalities such as Zoom in higher education. This study directly compares student conceptual learning outcomes in two parallel sections of an introductory physics course—one taught face-to-face (F2F) and the other synchronously via Zoom—delivered concurrently by the same instructor. Conceptual understanding of Newtonian mechanics was measured using the Force and Motion Conceptual Evaluation, administered presemester and postsemester. Normalized gain analysis showed average improvements of 30.49% for the F2F group and 25.01% for the Zoom group, a difference that was not statistically significant. Effect size analysis revealed a large gain for the F2F cohort ($d = 1.27$) and a moderate gain for the Zoom cohort ($d = 0.72$), with only a small, nonsignificant difference in post-test means ($d = 0.19$). Analysis of covariance confirmed that presemester scores were the strongest predictor of postsemester performance ($p < 0.001$), while modality itself was not significant ($p = 0.333$). A significant interaction ($p = 0.010$) indicated that prior preparation translated more effectively into learning gains for F2F students. Overall, both instructional modalities fostered meaningful learning; however, face-to-face instruction produced more consistent outcomes, whereas Zoom-based instruction revealed greater variability tied to students' initial levels of preparation.

DOI: [10.1103/j4st-rj18](https://doi.org/10.1103/j4st-rj18)

I. INTRODUCTION

The COVID-19 pandemic triggered a rapid and profound transformation in higher education, necessitating an almost overnight shift from traditional face-to-face (F2F) in-class instruction to various forms of online delivery. This unprecedented transition presented numerous global challenges for educators, learners, and institutions, including disruptions to curriculum, school closures, the urgent implementation of new teaching guidance, and often a lack of technological support and pedagogical content knowledge among university teachers for online platforms. This transition highlighted the potential of integrating online tools—such as synchronous video conferencing platforms like Zoom—into pedagogical practice [1].

While online learning systems like Mastering Physics, Expert TA, and WebAssign had existed prior to the pandemic, the entire educational community was abruptly thrust into an emergency remote teaching scenario [2]. Despite the widespread necessity of online instruction during the pandemic and the established efficacy of active

learning, a critical question regarding the specific modality of delivery remains: how does synchronous virtual instruction compare to traditional F2F in-class instruction when both are underpinned by consistent, evidence-based pedagogical strategies?

Prior research on online learning during the pandemic has yielded mixed or context-dependent results, often reflecting the challenges of emergency remote teaching rather than controlled comparisons. For instance, Ametepe and Khan [3] reported improved student grades in online physics courses during COVID-19 through strategies like recorded lectures, live demonstrations, and interactive opportunities. However, Pernia-Espinoza *et al.* [4] noted that the lack of hands-on activities in online STEM courses during lockdown correlated with less effective knowledge acquisition and diminished student attendance and participation.

Conversely, Cotter and Guldiken [5] successfully adapted active learning exercises for remote fluid mechanics courses via synchronous sessions, observing that while students initially preferred in-class learning, acceptance of online instruction grew overtime, though students consistently perceived a higher effort and responsibility online. Gamification, as explored by Rincon-Flores and Santos-Guevara [6], also showed promise in motivating engagement in online calculus courses. Furthermore, Klein *et al.* [2] found that students' self-organization and communication skills positively correlated with perceived learning achievement in online physics during the pandemic, and

*Contact author: ramesh.dhungana@ucdenver.edu

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.

that online labs were most effective when students could collect their own data.

While these studies offer valuable insights into various strategies and student perceptions in online environments, there is a distinct gap in direct, controlled comparisons of student conceptual learning outcomes between synchronous Zoom and traditional F2F in-class physics instruction when taught concurrently by the same instructor and with consistent, active engagement pedagogical strategies.

This study aims to fill this gap by presenting a direct comparison of student conceptual learning outcomes in two parallel sections of an introductory college physics course focused on Newtonian mechanics. One section was delivered via synchronous Zoom, and the other through traditional in-class, face-to-face instruction. Both sections were taught concurrently during a single postpandemic semester by the same instructor, and core instructional strategies and materials were designed to be consistent across both modalities.

Physics education has been a notable area of reform over the past three decades, driven by the growing field of Physics Education Research. PER aims to transform physics instruction and relies heavily on carefully designed research-based assessments to evaluate the effectiveness of teaching methods [7]. These assessments are particularly important for measuring conceptual understanding, as studies have shown that students can struggle with conceptual questions even when they perform well on quantitative problem-solving tests. Widely used conceptual assessments in mechanics include the Force Concept Inventory [8] and the Force and Motion Conceptual Evaluation (FMCE) [9]. The FMCE, specifically, is a research-based multiple-choice instrument designed to probe conceptual understanding of Newtonian mechanics.

To evaluate learning gains, this study employed both normalized gain and Cohen's d . While normalized gain is widely used in PER, recent research has identified its bias toward populations with higher pretest scores. Cohen's d , a standardized effect size metric, was included to provide a more statistically robust and equitable measure of learning gains across groups.

Furthermore, to statistically account for students' initial preparation or prior knowledge, an analysis of covariance (ANCOVA) was conducted using presemester assessment scores as a covariate. Initial analysis revealed that while regression lines for post-test scores vs pretest scores were approximately parallel across instructional modalities, the slopes differed significantly. This necessitated the inclusion of an interaction term in the ANCOVA model to accurately capture the differential impact of pretest scores on post-test outcomes between the F2F and Zoom groups.

The findings of this study contribute meaningfully to the ongoing discourse and research comparing Zoom-based virtual classrooms with traditional face-to-face instruction, offering insights into how delivery modality impacts

student learning. While the urgency of pandemic-era constraints may have subsided, this investigation remains relevant. Should another emergency arise that disrupts in-person education, educators can take comfort in knowing that synchronous virtual instruction can yield statistically comparable learning outcomes. Moreover, those currently offering or considering such courses will gain valuable insights into the effectiveness and practicality of Zoom as a viable instructional medium.

II. METHODS

A. Context and participants

This study was conducted at a public urban research university in the United States. The university has a substantial undergraduate population, with nearly 60% of its total enrollment comprising undergraduate students. The undergraduate student body is diverse, with approximately half identifying as students of color. The specific racial and ethnic composition beyond this broad categorization was not included in the scope of this study.

The participants in this study were undergraduate students enrolled in two parallel sections of an introductory college physics course focusing on Newtonian mechanics during the Fall of 2023 semester postpandemic. Most students enrolled in the Zoom section also participated in in-person laboratory sessions, indicating that they were not exclusively online learners. Based on observed engagement and feedback, it appears reasonable to conclude that these students preferred the Zoom format over traditional face-to-face instruction. Both sections were taught by the same instructor.

Sixty-four students were enrolled in the Zoom section of the course, which met on Mondays and Wednesdays at 8:00 AM. Two students withdrew during the semester. Of the remaining students, 47 completed both the presemester and postsemester surveys. Similarly, 57 students were enrolled in the face-to-face (F2F) section, held on Tuesdays and Thursdays at 10:00 AM. One student withdrew, and 46 completed both surveys. The surveys were administered during class via an online link. To encourage full participation, each student's lowest pre-class score and lowest weekly score were excluded from the grade calculation.

B. Instructional environment

The study involved two parallel sections of the introductory college physics course on Newtonian mechanics, taught concurrently by the same instructor during a single semester postpandemic. One section was delivered through a traditional in-class, face-to-face (F2F) modality, while the other utilized synchronous online instruction via the Zoom platform. The core instructional strategies and materials were designed to be consistent across both modalities to facilitate a direct comparison of learning outcomes based primarily on the delivery method.

The specific teaching methods utilized throughout the semester included:

1. Preclass preparation

Students were assigned preclass assignments administered via Canvas, the university's course management system. These assignments typically consisted of five knowledge-level multiple-choice questions. Students were provided with instructor-prepared PDF handouts attached to the assignments and were expected to read this material before coming to class.

2. In-class participation

During synchronous class sessions (both in-class and via Zoom), student participation was fostered through the use of iClicker response systems. These were used for administering conceptual questions and some problem-solving related questions.

3. Weekly assignments

Each week, students were assigned a worksheet prepared by the instructor and posted on Canvas. These worksheets contained a mix of problems and related tasks, such as drawing free-body diagrams and writing corresponding equations based on those diagrams.

4. Weekly tests

Weekly multiple-choice tests were administered via Canvas, consisting of three problem-related and two conceptual questions.

5. Major examinations

Two midterm exams and a cumulative final exam were administered to assess broader conceptual and problem-solving skills.

This combination of methods aimed to provide students in both the in-class and synchronous Zoom sections with opportunities for preclass preparation, active engagement during class time, and regular practice and assessment outside of formal lectures.

C. Presemester and postsemester assessments

To measure student learning over the duration of the semester, the FMCE was administered as both presemester and postsemester assessments. Only students who completed both assessments were included in the analysis.

The FMCE surveys were administered during class time using separate Qualtrics links for each section. Students were asked to enter their first and last names. They were informed that their lowest preclass quiz score would be dropped; however, participation was optional. Similarly, students were informed that their lowest weekly test score would be dropped if they participated in the postsemester

survey. Each class included a total of 11 weekly tests and 24 preclass quizzes during the semester indicating that there was only token amount of incentive to participate in the survey.

D. Data analysis

Student conceptual understanding of Newtonian mechanics was the primary outcome measured in this study. The dataset includes matched pretest and post-test scores for each student, along with a group identifier (0 = F2F, 1 = Zoom). The maximum possible score on the assessment was 47.

To quantify learning gains, we first calculated the normalized gain g for each section, defined as [10]:

$$g = \frac{\bar{x}_{\text{post}} - \bar{x}_{\text{pre}}}{\text{Maximum Score} - \bar{x}_{\text{pre}}}, \quad (1)$$

where, $\bar{x}_{\text{post}}$ and $\bar{x}_{\text{pre}}$ are average postsemester and presemester FMCE scores, respectively.

This metric represents the proportion of previously unknown material that a learner has acquired during the teaching program. While normalized gain is widely used in Physics Education Research (PER), recent literature [11,12] has identified its bias toward populations with higher pretest scores. Therefore, we also calculated Cohen's d . Cohen's d commonly used in other fields of social science that the Physics Education Research (PER) community has called for adoption [11].

Cohen's d for paired samples (e.g., pretest to post-test) is calculated by dividing the mean of the change scores for students by the standard deviation of those change scores. For independent samples, it divides the difference between the two sample means by the pooled standard deviation.

$$d = \frac{\bar{x}_{\text{post}} - \bar{x}_{\text{pre}}}{s_p}, \quad (2)$$

where s_p is pooled standard deviation of the pretest and post-test. It is given by

$$s_p = \sqrt{\frac{(n_{\text{pre}} - 1)s_{\text{pre}}^2 + (n_{\text{post}} - 1)s_{\text{post}}^2}{n_{\text{pre}} + n_{\text{post}} - 2}}. \quad (3)$$

For dependent samples, Cohen's d is calculated as [11]:

$$d_{\text{dep}} = \frac{\bar{x}_{\text{post}} - \bar{x}_{\text{pre}}}{\sqrt{s_{\text{pre}}^2 + s_{\text{post}}^2 - 2rs_{\text{pre}}s_{\text{post}}}} * \sqrt{2(1 - r)}, \quad (4)$$

where s_{pre} and s_{post} are the standard deviations and r is the correlation coefficients between pretest and post-test scores for each group.

Given that we have pre- and corresponding post-test data for each student, we calculate dependent Cohen's d_{dep} using Eq. (4) to compare pretest and post-test scores to

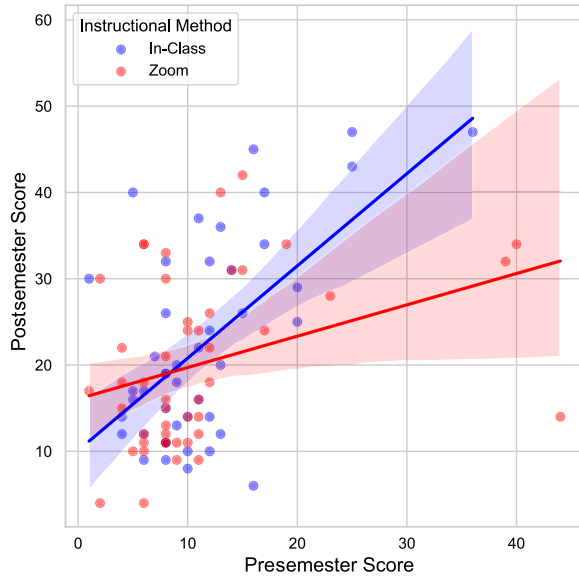


FIG. 1. Postsemester scores by instructional methods pretest scores as a covariate.

compute learning out of each group. We will calculate independent Cohen's d using Eq. (2) to compare the learning outcome between F2F and Zoom using corresponding post-test scores.

To further assess the impact of instructional modality while controlling for students' initial preparation, we conducted an analysis of covariance (ANCOVA). The model included pretest scores as a covariate and instructional modality as a fixed factor. Initial diagnostics revealed that the slopes differed significantly as seen in Fig. 1. This necessitated the inclusion of an interaction term between pretest score and instructional modality in the ANCOVA model. The model with the interaction term can be written as

$$\text{post} = \beta_0 + \beta_1 \text{pre} + \beta_2 \text{group} + \beta_3 (\text{pre} \times \text{group}), \quad (5)$$

where post and pre are post-test and pretest scores, respectively, group is delivery method (0 = F2F, 1 = Zoom). The coefficients β_0 , β_1 , β_2 , and β_3 and their values are discussed in the Results section. In Eq. (5), the last term is the interaction term, which captures whether the relationship between pre and post scores differs by group (e.g., F2F vs Zoom).

The adjusted ANCOVA model allowed us to evaluate whether the relationship between pretest and post-test scores varied by instructional format. This approach supports a more nuanced interpretation of equity of opportunity and instructional effectiveness [12].

III. RESULTS

This section presents the results of the comparative analysis of student learning outcomes between the

TABLE I. Descriptive statistics for each group.

Statistics	F2F $\pm$ SD	Zoom $\pm$ SD
Number of students (N)	[46]	[47]
Average pretest score	[11.28 $\pm$ 6.38]	[11.15 $\pm$ 8.99]
Average post-test score	[22.17 $\pm$ 11.64]	[20.13 $\pm$ 9.80]
Average normalized gain	[30.49%]	[25.01%]

face-to-face (F2F) and synchronous Zoom sections using three complementary approaches: normalized gain, Cohen's d , and analysis of covariance (ANCOVA) with interaction.

A. Normalized gain

Normalized gain (g) was calculated using the average pretest and post-test scores for each group, following the traditional method in Physics Education Research (PER) [10], using the formula in Eq. (1):

Table I provides a summary of the descriptive statistics for each group, including the number of students included in the analysis (N), the average pretest and post-test scores with standard deviations, and the calculated average normalized learning gain.

Figure 2 visualizes average normalized gain with standard deviations.

These results suggest that students in the F2F section achieved slightly higher normalized gains than those in the Zoom section. However, the difference in average normalized learning gains between the in-class and synchronous Zoom sections was not statistically significant.

However, as noted by Nissen *et al.* [11], normalized gain can be biased toward populations with higher pretest scores and may not be suitable for equity comparisons. We also calculated Cohen's d . See Sec. II D for more details.

B. Cohen's d effect sizes

To provide a more statistically robust measure of learning gains, Cohen's d was calculated for both within-group (pre vs post) and between-group (post-test) comparisons [11,12]. Cohen's d is a widely used measure of

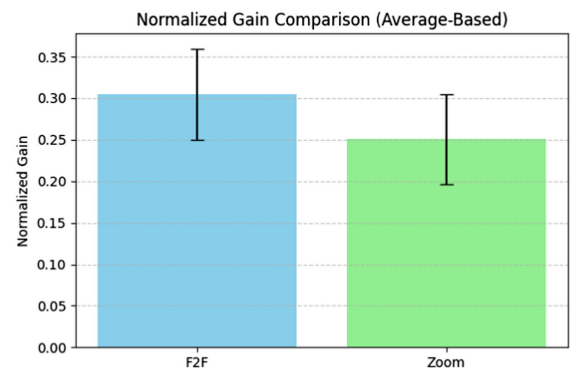


FIG. 2. Normalized gain with uncertainty.

TABLE II. Conventional thresholds for interpreting Cohen's d [13].

Effect size	Range of d	Interpretation
Small	$0.2 \leq d < 0.5$	Subtle but potentially meaningful
Medium	$0.5 \leq d < 0.8$	Moderate visible to the naked eye
Large	$d \geq 0.8$	Substantial practical impact

TABLE III. Descriptive statistics and effect sizes for F2F and Zoom sections.

Statistic	F2F	Zoom
Mean pretest	11.28	11.15
Mean post-test	22.17	20.13
SD pretest	6.38	8.99
SD post-test	11.64	9.80
Correlation (Pre-post)	0.586	0.333
Cohen's d (dependent)	1.27	0.72
Cohen's d (independent, post F2F vs Zoom)	0.19	

standardized effect size, particularly in behavioral and educational research. According to Cohen's conventional benchmarks [13], effect sizes can be interpreted as shown in the Table II:

In the present study shown in Table III, the face-to-face (F2F) section exhibited a large effect size ($d = 1.27$), indicating substantial conceptual gains. The Zoom section showed a medium effect ($d = 0.72$), suggesting meaningful but less pronounced improvement. The between-group comparison of postinstruction scores yielded a small effect size ($d = 0.19$), which was not statistically significant.

These results indicate that both groups experienced large learning gains, with the F2F group showing slightly higher improvement. The small effect size in the post-test comparison suggests that final performance levels were relatively similar.

C. ANCOVA with interaction term

To assess the impact of instructional modality while controlling for prior knowledge, an ANCOVA model was fitted with pretest score as a covariate and instructional group as a factor. Given that the regression lines for post-test vs pretest scores were not parallel as seen in Fig. 1, the Eq. (5) with interaction term was used:

$$\text{post} = \beta_0 + \beta_1 \cdot \text{pre} + \beta_2 \cdot \text{group} + \beta_3 \cdot (\text{pre} \times \text{group}),$$

where:

- $\beta_0 = 10.112$: Estimated post-test score for F2F students with assumed pretest score of 0.
- $\beta_1 = 1.069$: For F2F students, each additional point on the pretest increases the post-test score by approximately 1.07 points.

TABLE IV. ANCOVA model with interaction term.

Term	Coefficient	p value
Intercept	10.112	0.001
Pretest	1.069	<0.001
Group (Zoom)	5.967	0.102
Pre $\times$ group	-0.706	0.010

- $\beta_2 = 5.967$: Zoom students score approximately 6 points higher than F2F students when pretest score is 0.
- $\beta_3 = -0.706$: Zoom students gain approximately 0.71 fewer post-test points per pretest point compared to F2F students.

The model summary is shown in Table IV. The interaction term, $\beta_3 = -0.71$, was statistically significant ($p < 0.05$), indicating that the relationship between pretest and post-test scores differed by instructional modality. This violates the homogeneity of regression slopes assumption and justifies the inclusion of the interaction term [12].

These findings suggest that the instructional modality moderated the effect of prior knowledge on post-test performance, with the F2F group benefiting more from higher pretest scores.

IV. CONCLUSION

This study examined the impact of instructional modality on student learning in an introductory physics course, comparing traditional face-to-face (F2F) instruction with synchronous online teaching via Zoom. Student learning outcomes were evaluated using normalized gains, effect sizes (Cohen's d), and analysis of covariance (ANCOVA) with presemester scores as a covariate, providing complementary perspectives on achievement and consistency.

Both instructional modalities yielded significant improvements from preinstruction to postinstruction, as shown in Table I. The F2F group increased from an average of 11.28 ± 6.38 to 22.17 ± 11.64 , while the Zoom group improved from 11.15 ± 8.99 to 20.13 ± 9.80 . The average normalized gains were 30.49% for F2F and 25.01% for Zoom, a modest difference that was not statistically significant, indicating that overall rates of improvement relative to starting knowledge were comparable between modalities.

By contrast, Cohen's d effect sizes revealed differences in the magnitude and consistency of learning gains: the F2F cohort achieved a large effect size ($d = 1.27$), while the Zoom cohort reached only a moderate level ($d = 0.72$). A direct comparison of post-test means yielded only a small and nonsignificant effect ($d = 0.19$), suggesting that average outcomes were broadly similar despite variability in effect sizes.

The ANCOVA analysis provided further nuance. Presemester scores were a strong predictor of postsemester

outcomes ($\beta \approx 0.60$, $p < 0.001$), while modality alone was not a significant predictor ($\beta = -1.97$, $p = 0.333$). Importantly, the significant interaction between pretest score and modality ($\beta = -0.71$, $p = 0.010$) indicated that prior preparation translated more strongly into gains for F2F students than for Zoom students. This interaction explains why the F2F group exhibited greater consistency, while the Zoom group showed wider variability in outcomes.

Taken together, these findings highlight a key contrast. When judged by normalized gains, F2F and Zoom instruction appear equally effective on average. However, effect size and ANCOVA analyses reveal that F2F instruction produces more consistent outcomes and amplifies the benefits of higher prior preparation, whereas Zoom instruction yields greater variability across students. This divergence underscores that instructional modality influences not just mean achievement, but also the translation of prior knowledge into postinstruction performance.

In conclusion, synchronous online instruction via Zoom can support meaningful learning gains in introductory

physics, but it may not fully replicate the consistency observed in traditional classrooms especially for those with higher pretest scores. Future research should investigate how instructional design, engagement strategies, and systemic supports can reduce variability in online settings and promote equitable learning opportunities across modalities.

ACKNOWLEDGMENTS

We thank the students who participated in this study. We also gratefully acknowledge the support of the Department of Physics at the University of Colorado Denver. We thank one anonymous reviewer for their thoughtful suggestions on improving the presentation of this manuscript.

DATA AVAILABILITY

The data that support the findings of this article are openly available [14].

-
- [1] L. R. Krome, A review of Zoom utilization in higher education during the COVID-19 pandemic, *Educ. Thinking* **1**, 11 (2020), <https://analyttrics.org/article/a-review-of-zoom-utilization-in-higher-education-during-the-covid-19-pandemic/>.
 - [2] 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).
 - [3] J. D. Ametepe and N. Khan, Teaching physics during COVID-19 pandemic: Implementation and report of teaching strategies to support student learning, *Phys. Educ.* **56**, 065030 (2021).
 - [4] A. Pernia-Espinoza, A. Sanz-Garcia, F. J. Martinez-de Pison-Ascacibar, and F. Navaridas-Nalda, Active learning methodologies in STEM degrees jeopardized by COVID-19, in *Proceedings of the 2021 IEEE Global Engineering Education Conference (EDUCON)*, Vienna, Austria (IEEE, New York, 2021), pp. 689–695.
 - [5] J. M. Cotter and R. Guldiken, Remote versus in-class active learning exercises for an undergraduate course in fluid mechanics, in *Proceedings of the 2021 ASEE Virtual Annual Conference* (2021), <https://peer.asee.org/37661>.
 - [6] E. G. Rincon-Flores and B. N. Santos-Guevara, Gamification during COVID-19: Promoting active learning and motivation in higher education, *Australas. J. Educ. Tech.* **37**, 43 (2021).
 - [7] J. Von Korff, B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell, Secondary analysis of teaching methods in introductory physics: A 50 k–student study, *Am. J. Phys.* **84**, 969 (2016).
 - [8] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
 - [9] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton’s laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
 - [10] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
 - [11] J. M. Nissen, R. M. Talbot, A. N. Thompson, and B. Van Dusen, Comparison of normalized gain and Cohen’s d for analyzing gains on concept inventories, *Phys. Rev. Phys. Educ. Res.* **14**, 010115 (2018).
 - [12] E. Burkholder, C. Walsh, and N. G. Holmes, Examination of quantitative methods for analyzing data from concept inventories, *Phys. Rev. Phys. Educ. Res.* **16**, 010141 (2020).
 - [13] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge, New York, 1988).
 - [14] R. Dhungana, Pre-post data of zoom and in-class sections (2025), [10.7910/DVN/B6OWRV](https://doi.org/10.7910/DVN/B6OWRV).