Pre-lab video boosts Z-score understanding in introductory physics labs

Tianlong Zu & Aakash Lakshmanan
Department of Physics and Astronomy, Northwestern University,
2145 Sheridan Road, Evanston, IL, USA 60208

Introductory physics labs have long been a cornerstone of undergraduate physics education, owing to their integral role in the curriculum. However, students often encounter challenges, particularly in data analysis. This includes grappling with various concepts and methodologies, with uncertainty analysis, notably the Z-score, standing out as a persistent hurdle. In our study, we aimed to address this issue by integrating a concise instructional video, aligning with the Cognitive Theory of Multimedia Learning principles, into the initial stage of an introductory physics lab focusing on data analysis. In the study, ninety-four students were exposed to and engaged with the Z-score video, while ninety-one students did not. The outcomes unveiled a notable divergence in performance between the video and non-video groups, evident both immediately and after a week. These results strongly affirm the potency of concise multimedia interventions in augmenting student understanding and proficiency in introductory lab instruction. Limitations and future directions were also discussed in the end.

2024 PERC Proceedings edited by Ryan, Pawl, and Zwolak; Peer-reviewed, doi.org/10.1119/perc.2024.pr.Zu

I. INTRODUCTION

A popular aim in introductory physics labs and one emphasized by the AAPT [1] is that of reinforcing, in lab, lecture concepts. The effectiveness of such an aim however has been questioned in recent years appealing to the fact that they encourage less expert-like epistemologies in their understanding of how theory relates to experiment [2], another important goal set forth by the AAPT. One epistemological survey [3] noted that students believe that the sole criterion of validity for their experimental procedures is agreement with theory. In contrast, PhD students often consider a variety of experimentally intrinsic factors, such as the quality of work, uncertainty analysis, and repeatability, when conducting experiments. “Cookbook”-style replication of lecture results only serves to reinforce such misconceptions.

A possible way to approach this is to look at how students understand uncertainty analysis, another goal set forth by the AAPT. In the same epistemological survey mentioned above, students reported that they see uncertainty analysis as solely a means of checking with theory. This is due to students subscribing to what is known as the "point paradigm," which assumes that measurement is about singular values whose distance from the "real" value is a metric of imperfection and thus the goal is to reduce it to 0. This paradigm disproportionately focuses on accuracy rather than precision, a crucial aspect of scientific measurements [4] and ignores the importance of many measurements. An alternative paradigm is the "set paradigm," which includes uncertainties in the value itself, emphasizing the collective nature of measurement. The goal is that students would value the importance of uncertainty and start paying attention to multiple measurement and checking the consistency of measurements.

However, internalizing this understanding of set paradigm has proven to be challenging [5, 6]. Students could internalize the paradigm through the use of Z-score, defined as[1]:

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\sigma_{E1}^2 + \sigma_{E2}^2}} \qquad (1)$$

Where $\bar{x}_1$ and $\bar{x}_2$ represent the *sample mean values*. σ_{E1} and σ_{E2} represent the *standard deviation of the means*. Equation (1) assumes a vanishing difference between the two population means. σ_{E1} and σ_{E2} are estimated using the equation $\sigma_E = sample\ S.D./\sqrt{sample\ size}$.

This z-score formula is employed to compare two mean values from two normally distributed but independent populations. This is particularly opposed to looking at range overlap or percent differences which still push toward point paradigms. Holmes and her collaborators have shown that emphasizing and reiterating this metric has significant impacts on experimental method, including more trials conducted, greater precision attained, and more nuanced reflection of the results [4]. Students were even confident enough to disagree with standard theory, such as with the periods of pendula with high amplitude.

It is upon these considerations that we chose to emphasize Z-score concept which both directly and indirectly accomplishes many of the goals we hope to achieve in a laboratory environment.

The core of our approach is embedded video designed with multimedia learning principles which have been shown to be effective and resource-efficient in other learning environments. For example, the pre-lecture multimedia learning modules (MLMs) at the University of Illinois at Urbana-Champaign have been shown to produce very promising results in the lecture room. These studies along with others have shown that the MLMs increased student performance [7-9], engagement [10], and value gained from the lecture, while decreasing perceived difficulty [11]. It also rests on firm principles set by cognitive load theory and Mayer’s Cognitive Theory of Multimedia Learning principles [12] which are elaborated below.

We propose to test the effectiveness of a pre-lab video intervention on understanding of Z-score concept and measurement in the context of introductory physics labs. We will demonstrate that such intervention has shown some initial evidence of effectiveness in both the short and long term while providing minimal cost in resources.

II. METHODOLOGY

Participants comprised of undergraduates concurrently enrolled in the lab course in an elite private university in the Midwest. Most were taking calculus-based physics and were science and engineering majors, but a few were taking algebra-based physics courses. Many students also took one or two prerequisite physics lab courses prior to the one during this study.

[1] Please note that the discussion here follows OpenStax text Introductory Statistics 2e Chap 10.2: Two Population Means with Known Standard Deviations, where the general formula for z-score is:

$$z = \frac{|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)|}{\sqrt{\sigma_{E1}^2 + \sigma_{E2}^2}}$$

Please note the notation difference here. We again used σ_{E1} and σ_{E2} to represent the *standard deviation of the means*, such that the sample sizes were not explicitly shown in the equation as compared to the formula in the OpenStax text. We also assumed $\mu_1 - \mu_2 = 0$, since it is typical to adopt two different methods to measure the same physics quantity in an introductory physics lab. In this case, the expected different in population means should vanish.

In the case of a single population, the z-score formula for a normally distributed random variable is:

$$z = \frac{x - \mu}{\sigma}$$

However, it is rarely adopted in introductory lab instruction since it deals with a single measurement rather than multiple measurements.

The lab course they were enrolled in for this study is the third in the introductory lab sequence which was primarily focused on waves, optics, and modern physics. Students engage in guided worksheets that take them through in-person labs or simulations or in order to confirm and check the consistency between various measurements of some quantity e.g. wave speed, focal length, index of refraction, etc. The labs are taught in a format closer to the traditional method than inquiry based, design oriented, or exploration focused style. The lab that occurred during this study specifically was the third, an introduction to data analysis where they filled in a worksheet to learn about properties of measurement tools (digital and analog), measurement in physics, computation of averages, S.D., S.E., and Z-scores.

A total of 185 students from 10 sections were selected for this study. We followed the common practice of convenient sampling such that 5 sections were randomly assigned to group 1 and the other 5 group 2 resulting in N=94 in group 1 and N=91 in group 2. Group 1 watched the Z-score video before the lab started. Both groups were given a pre-test, assigned the lab, a post-test and then a delayed test in the beginning of the next lab which is one week later. The same test was used for the posttest and the delayed test.

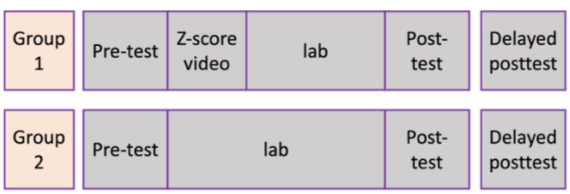

FIG.1 experimental procedure of the study

The Z-score video[2] (a total duration of 5 mins & 31 secs) was created based on Mayer's Cognitive Theory of Multimedia Learning principles. Mayer's CTML principles rest on a few basic assumptions: dual channeling i.e. students possess an auditory and a visual channel of information processing, limited cognitive capacity i.e. students have a limit on information able to be processed simultaneously, and active processing i.e. students are actively constructing mental representations. These translate to practical principles such as using simple narration and images without interfering text, no extraneous material, having clear centers of attention, using informal voice, making voiceover and images simultaneous, and dividing the video into small chunks. We have this video checked by two experts in the field and both agree that the content and delivery method were proper for this population, and the design consistent with Mayer's principles.

The *questions on the test* are accessible following this link[3]. The multiple-choice questions were designed to probe a critical understanding of Z-score concepts. The free response question was designed to understand how students would approach a standard measurement as a probe of experimental measurement, data interpretation, and Z-score.

III. RESULTS

We first introduce the coding and scoring scheme for analyzing student answers to the free response question. Subsequently, we integrated these scores with those from multiple-choice questions (MCQs) and proceeded with comprehensive analysis of the aggregate score.

A. Coding analysis for the free response question

We followed an emergent coding approach to analyze student answers to the free response question. Two independent coders analyzed 10% of the data separately and produced different codes. Initially, there was disagreement on the number of codes and 2 more codes were created to accommodate. Following the new code scheme, a different 10% of the data was coded with an interrater reliability of 0.74 on the pretests, 0.79 on the post tests, and 0.79 on the delayed posttests as evaluated by Cohen's κ. After conferring, this reached 1.0. The codes are given in Table I.

Code (score assigned)	Criteria
Code 1: *Single measurement* (1 point)	Makes mention of tool, consideration of accuracy (e.g. posture)
Code 2: *Measurement tool property* (1 point)	Mentions resolution of the measurement tool used
Code 3: *Multiple measurements* (2 point)	Mentions multiple measurements
Code 4: *Uncertainty* (2 point)	Mentions reporting uncertainty, checking for agreement/consistency

TABLE I: Code for the free response question

Code 1 *single measurement* describes students taking note of one-time measurement and variables like posture, parallax, etc. which are certainly related to the measurement of a person's height as in the question. However, it does not differ greatly from the answer of an every-day person who had not taken the physics lab. This was awarded a score of 1 point.

Code 2 m*easurement tool property* describes students expressing the measurement tool should have high resolution. This was awarded a score of 1 point.

Code 3: *Multiple measurements* speak to measuring the height multiple times and select the mean to best represent the height. This code is consistent with how physicists

[2] Readers can watch this video following this link: https://drive.google.com/file/d/14U3UR6r2eAtXMtmGg0EnpfcXbZMtFtp6/view?usp=sharing

[3] https://drive.google.com/file/d/1-Ji-1FzqAwFiyNUDKfMQTFJWBzq92Fpn/view?usp=sharing Please note the yellow color highlighted options are the expected answers.

conduct experiments and is what we want students to learn. We rewarded it with a score of 2 points.

Code 4: *Uncertainty* discusses calculating uncertainty after multiple measurement and many students expressed making the uncertainty as small as possible. Some students also mentioned using uncertainty to obtain a z-score. This code is consistent with how physicists conduct experiments and is what we want students to learn. We assigned a score of 2 points to this code.

We were particularly interested in seeing if students had developed the "set paradigm" i.e. that multiple measurements and reporting uncertainty and consistency is always vital. Additionally, many students noted multiple measurements but far fewer noted checking consistency or reporting uncertainty. Example answers and assigned codes are given in Tables II.

Example excerpt 1: "I would have the person remove their shoes and socks and stand straight against a wall. Ideally the person would be bald so the hair doesn't add height. I would then take a ruler or some other straight rod and place it at the top of their head at a 90 degree angle to the wall. Then I would take a tape measure and measure the length from the floor to the bottom of the ruler to measure their height." **Coding**: An everyday answer that makes considerations of accuracy with taking shoes and socks off (**code 1**) **Total score**: 1 point
Example excerpt 2: "In order to measure something reliably you need to ensure that all values are being held constant. We want to make sure that our subject is standing tall, with no knees bent or shoulders shrugged. We should mark their stance as we will want to take multiple measurements and average them. In addition to taking multiple measurements, we need to account for the uncertainty. We should calculate a standard deviation as well as a z-score, to have the most reliable measurements. We should be using the highest precision measuring devices that we have availability to in calculating their height." **Coding:** Considerations of accuracy with posture (**code 1**), high precision tools (**code 2**) multiple measurements (**code 3**), and uncertainty/Z-score (**code 4**). **Total score:** 6 points

Table II: Examples of coded answers

B. Total score analysis

Based on the rubric laid out in the above section, we assigned scores to each student and add that to their MCQ scores to determine a total score for each student. We examined students' total score since Q1 probes students' understanding of means, uncertainty, etc. which would lead to z-score which is the focus of Q2-4. Each MCQ was rewarded 1 point if the expected answer was selected. Statistics of the total scores for the two groups can be found in Table III. To determine what analysis was the most appropriate, we first conducted a Shapiro-Wilk normality test on students' total scores in the pretest, posttest, and the delayed test. We found that the assumption of normality is violated in the pretest ($W = 0.8548, p < 0.001$), in the posttest ($W = 0.8581, p < 0.001$), and in the delayed test ($W = 0.8116, p < 0.001$).

Total	Pretest	posttest	Delayed test
Group 1	2.51 $\pm$0.17	4.02$\pm$0.19	4.38$\pm$0.21
Group 2	1.85$\pm$0.17	3.27$\pm$0.22	3.12$\pm$0.22

Table IV: Means and standard errors of the total score

Since the normality assumption was violated in the pretest, Mann-Whitney U test (for large sample size) was conducted to determine whether there is a difference in the pretest total scores between the two groups The results indicate a significant difference, $W = 5199.5, p = 0.007$. A boxplot of the pretest total scores for the two groups was presented in Figure 2.

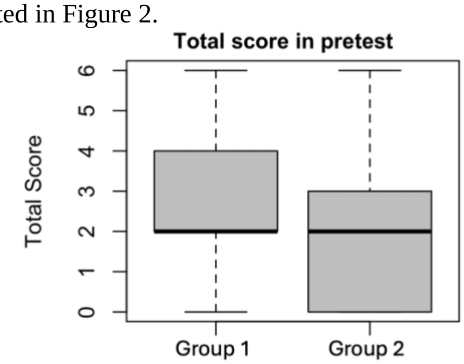

FIG.2 Box plot of student performance on the free response question on the pretest between the two groups

Building upon this notable finding, we employed the pretest score as a covariate in a subsequent ANCOVA to compare students' overall performance scores in the posttest and the delayed test. Our analysis revealed a significant discrepancy in performance on the total scores in the posttest, even after accounting for the influence of pretest performance, $F(1, 182) = 4.1252, p = 0.044$, Specifically, students who viewed the z-score video exhibited notably superior performance compared to those who did not, a trend that persisted even after controlling for pretest performance.

We conducted a similar ANCOVA to assess student performance on the overall survey in the delayed test, employing pretest performance on the same questions as a covariate. The results unveiled a notable between-group disparity, $F(1, 182) = 11.391, p < 0.001$, indicating that students who engaged with the z-score video showcased markedly superior performance compared to their counterparts who did not, a trend that persisted even after adjusting for pretest performance.

IV. DISCUSSION

In this study, we aimed to assess the impact of incorporating a concise pre-lab video on introductory physics students' grasp of experimental measurement and uncertainty analysis, specifically focusing on the Z-score concept. Utilizing emergent coding analysis, we distilled four representative codes for scrutinizing student responses to the free-response question. Examination of students' total scores revealed a significant advantage for those who viewed the videos over those who did not, both in the posttest and the delayed test, even after accounting for pretest disparities. These findings underscore the tangible benefits of pre-lab video supplementation, suggesting enduring improvements in comprehension over extended periods, such as a week.

Students who spent time watching the video showed significant performance on the posttest and delayed test. This result provides some initial evidence that this student group demonstrated a better shift to the set paradigm indicating more awareness of the significant of multiple measurement and checking for consistency using standard deviation, and z-score.

Given many of these students had already been introduced to these ideas in previous labs, it is remarkable that such a small intervention still makes such drastic improvement. This begs the question of why such a phenomenon appears. We propose that not only does it serve as a reminder of concepts, but it explains the idea in detail whereas teaching assistants often focus on simply knowing how to compute the necessary values. This causes better understanding and retention. This strategy could be interest to many educators since it could well supplement many TAs' instruction which is a typical practice in many physics departments. Creating videos for various important concepts in the course may be an important step forward.

Even though this study demonstrates even a small intervention might help getting students started on the shift from point paradigm to set paradigm, limitations of the study exist. First of all, the intervention was conducted only in one session of the whole lab course, future studies should look at the effect of implementing short videos in the beginning of each lab session for a whole term. Secondly, we chose the common practice of convenient sampling in the study design, future studies should use authentic random sampling in the design to see if the results hold. Lastly, the study involved students from only one institution, future studies should look at if the results can be replicated in different types of institutions.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation grant 2111138. Opinions expressed are of the authors and not necessarily of the Foundation. We would like to thank all the TAs and students who participated in the study.

[1] https://www.aapt.org/resources/policy/goaloflabs.cfm
[2] N. G. Holmes, J. Olsen, J. L. Thomas, and C. E. Wieman, Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content, Phys. Rev. Phys. Educ. Res. 13, 010129 (2017).
[3] D. Hu and B. M. Zwickl, Examining students' views about validity of experiments: From introductory to Ph.D. students, Phys. Rev. Phys. Educ. Res. 14, 010121 (2018).
[4] N. G. Holmes and D. A. Bonn, Quantitative comparisons to promote inquiry in the introductory physics lab, Phys. Teach. 53, 352 (2015).
[5] T. S. Volkwyn, S. Allie, A. Buffler, and F. Lubben, Impact of a conventional introductory laboratory course on the understanding of measurement, Phys. Rev. ST Phys. Educ. Res. 4, 010108 (2008).
[6] R. L. Kung and C. Linder, University students' ideas about data processing and data comparison in a physics laboratory course, Nordic studies in science education, 2, 40 (2006).
[7] T. Stelzer, G. Gladding, J. Mestre, and D. T. Brooks, Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content, Am. J. Phys. 77, 184 (2009).
[8] Z. Chen, T. Stelzer, and G. Gladding, Using multimedia modules to better prepare students for introductory physics lecture, Phys. Rev. ST Phys. Educ. Res. 6, 010108 (2010).
[9] M. Hill, M. D. Sharma, and H. Johnston, How online learning modules can improve the representational fluency and conceptual understanding of university physics students, Eur. J. Phys. 36, 045019 (2015).
[10] H. R. Sadaghiani, Using multimedia learning modules in a hybrid-online course in electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. 7, 010102 (2011).
[11] T. Stelzer, D. R. Brookes, G. Gladding, and J. P. Mestre, Impact of multimedia learning modules on an introductory course on electricity and magnetism, Am. J. Phys. 78, 755 (2010).
[12] R. E. Mayer, Cognitive theory of multimedia learning, in The Cambridge Handbook of Multimedia Learning (Cambridge University Press, New York, 2005).
[13] 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).