Comparison of expert-like attitudes and scientific reasoning skills

Amber Sammons
South Fork Jr/Sr High School, Kincaid, IL 62540

Rebecca Rosenblatt
National Science Foundation EHR/HRD Alexandria, VA 22314
Please note that the opinions, findings, conclusions and recommendations expressed here
are the authors' and do not necessarily reflect the views of the National Science Foundation

Raymond Zich
Department of Physics, Illinois State University, Normal, IL 61790

Results from a multi-semester study of the effects of eight supplemental laboratory activities in a general education physics course will be presented. A total of two control and three treatment semesters were studied. The results allowed comparison between expert-like attitudes measured by the Colorado Learning Attitudes about Science Survey (CLASS) and scientific reasoning skills measured by Lawson's Classroom Test of Science Reasoning. Correlation of the pre/posttest CLASS scores and posttest Lawson scores found no relationship between the scores. Both student attitudes and scientific reasoning skills showed improvement, relative to a control semester, for the first semester the intervention was applied. In subsequent semesters, improved scientific reasoning skills continued to be observed, but not improvement in students' scientific attitudes. A detailed comparison of the CLASS and Lawson scores are presented along with a discussion of implications for instruction given this apparent decoupling of expert-like attitudes and reasoning skills.

I. INTRODUCTION

This research presents results from a five-semester study investigating the interplay between students' scientific reasoning skills and their attitudes about science. Results show that students' scientific reasoning skills, as measured by posttest with the Lawson Classroom Test of Scientific Reasoning (Lawson) [1], were stable within condition and, assuming equal pretest scores, suggests improvement with the use of an intervention curriculum. On the other hand, students' attitudes about science, as measured by the Colorado Learning Attitudes about Science Survey (CLASS) [2], showed large variations across semesters and were not consistently improved by the intervention. In addition, correlations between scores of students' reasoning skills and their attitudes revealed that overall these are only weakly correlated for this student population, which suggests expert-like attitudes are not necessary for the development of reasoning.

In physics courses, and scientific education in general, a major goal in addition to acquiring content knowledge, is the attainment of scientific reasoning skills. Instructors strive to develop students' capabilities to follow the processes and products of scientific research, but also work to develop students' deeper understanding of the logical methodologies that underlie scientific investigation. These scientific reasoning skills are closely connected to formal operant reasoning and critical thinking skills [3, 4] and are the skills used during inquiry, experimentation, evaluation of evidence, and inference [5, 6]. In other words, these are the skills that allow students to "do science". In addition, there has been increased understanding that students' attitudes, beliefs, and epistemologies about physics play a significant role in learning and retention [7–9]. This has lead to the development of instruments such as the Maryland Physics Expectations Survey (MPEX) [10], Views About Science Survey (VASS) [11], Epistemological Beliefs Assessment for Physical Sciences (EBAPS) [12], and the CLASS [2] to measure these interrelated attitudes, beliefs, and epistemologies.

While it is clear that the areas of scientific reasoning and attitudes toward science are individually well-studied, there has been very little work that investigates how, if at all, these two important areas are related. None of the papers citing the CLASS in the PER Central database, including the larger 2015 meta-analysis of student attitude research [9], investigated the relationship between students' attitudes as measured by these instruments and their skills in scientific reasoning. Attitudes have been contrasted with other measures of physics content knowledge such as FCI and FMCE [8, 13], but not with reasoning itself. In addition, the literature tends to use nebulous language that groups physics content skills together. Examples of this vague language are seen in the phrases "significant conceptual gains were found using a conceptual instrument" [13] and "significant improvements in both content knowledge and beliefs" [14].

In this research study, four main research questions exploring the relationship between attitudes and science reasoning were investigated: 1. Does the curriculum created improve students' scientific reasoning skills compared with the original course?; 2. Does the curriculum created result in more expert-like student attitudes?; 3. Are attitudes, expert-like or novice-like, related to scientific reasoning? (This was operationalized into the question: How does student performance on the CLASS correlate with student performance on the Lawson test? i.e. Do students need to think like physicists to be able to reason scientifically?); 4. Are certain categories of attitudes, such as problem solving sophistication or personal interest, more highly correlated with reasoning expertise?

A. The Colorado Learning Attitudes Science Survey (CLASS)

The CLASS is a 5-point Likert scale assessment. It uses 42 statements to gauge each student's attitude about science as being expert-like or novice-like depending on if they agree/disagree with the statements. Of the 42 statements, 18 are statements experts disagree with (having a lower score on the scale), 18 are statements experts agree with (having a higher score on the scale), and 6 are statements that are neutral or with no clear expert consensus. The test has no traditionally right or wrong answers and is scored on whether students agree with expert views or novice views, which is consistent with many other attitude assessments [7]. The CLASS is often administered as a pretest and posttest to judge students' shift in attitudes across the semester [2, 8, 13–20].

The question statements are grouped into eight categories that measure different aspects of students' attitudes about science. These categories are: personal interest, real world connection, problem solving general, problem solving confidence, problem solving sophistication, sense making/effort, conceptual understanding, and applied conceptual understanding (see Table I). It should be noted that some questions appear in more than one category. The CLASS test has been validated [2] and is commonly used as a method for evaluating a curriculum's effect on students' attitudes about science [8, 13–20]. While a majority of these studies indicate reliable and consistent results, as documented in the meta-analysis by Madsen et. al. [9], there is some evidence of imperfections in the CLASS [21].

TABLE I. Categories for CLASS Questions.

Categories	Questions
Personal Interest	3, 11, 14, 25, 28, 30
Real World Connection	28, 30, 35, 37
Problem Solving General	13, 15, 16, 25, 26, 34, 40, 42
Problem Solving Confidence	15, 16, 34, 40
Problem Solving Sophistication	5, 21, 22, 25, 34, 40
Sense Making/Effort	11, 23, 24, 32, 36, 39, 42
Conceptual Understanding	1, 5, 6, 13, 21, 32
Applied Conceptual Understanding	1, 5, 6, 8, 21, 22, 40

TABLE II. Scores for each semester with (standard deviation). CLASS is scored by percentage of expert-like responses on the agree/disagree questions. Cor.[Pre, Post, Δ]-C/Law. are the correlations between score on the respective CLASS test and the Lawson test. Δ is the Change.

Semester	N	Lawson	Pre-CLASS	Post-CLASS	Change CLASS	Cor.Pre-C/Law.	Cor.Post-C /Law.	Cor.Δ-C/Law.
Control 1	93	54% ($\pm$18%)	48% ($\pm$15%)	38% ($\pm$17%)	-10% ($\pm$16%)	0.210	0.166	-0.018
Control 2	41	57% ($\pm$19%)	42% ($\pm$13%)	44% ($\pm$15%)	+2% ($\pm$15%)	0.092	0.020	0.189
Control Average	134	55% ($\pm$19%)	46% ($\pm$15%)	40% ($\pm$17%)	-6% ($\pm$16%)	0.157	0.139	0.055
Experiment 1	48	58% ($\pm$18%)	49% ($\pm$19%)	44% ($\pm$20%)	-5% ($\pm$16%)	0.293	0.361	0.098
Experiment 2	46	64% ($\pm$18%)	61% ($\pm$16%)	34% ($\pm$12%)	-27% ($\pm$15%)	0.429	0.300	-0.252
Experiment 3	22	61% ($\pm$21%)	44% ($\pm$14%)	50% ($\pm$12%)	+6% ($\pm$15%)	-0.201	-0.571	-0.240
Experiment Average	116	61% ($\pm$18%)	52% ($\pm$18%)	41% ($\pm$17%)	-10% ($\pm$19%)	0.254	0.127	-0.141

B. The Lawson Test of Scientific Reasoning

The Lawson test was developed from the Lawson Classroom Test of Formal Reasoning [1, 22]. There are six categories covering the reasoning areas of conservation, proportional reasoning, control of variables, probability, correlational reasoning, and hypothetical-deductive reasoning [1, 22]. It is a two-tier multiple-choice (TTMC) test with 12 pairs of questions. The first question of the pair asks for the correct response for a given situation, and the second question asks for the explanation or reasoning behind the answer to the first question. There are two ways to score a TTMC. The pair scoring method assigns one point for answering both questions in a pair correctly, and the individual scoring method that scores each question independently. Both methods have been used in PER [1, 5, 22, 23]. In this study, the individual scoring method was used. The Lawson test has been validated [22] and is commonly used in research studies [3, 5, 6, 18].

II. METHODOLOGY

Data was collected in multiple semesters of an algebra-based general-education physics course for non-science majors at a large public primarily undergraduate institution. The class was taught in a large lecture format that met two or three times a week, depending on the semester, and had a separate weekly lab. A total of two control (N = 134) and three treatment (N = 116) semesters were studied with the same instructor teaching each semester. Students were given the CLASS test as a pretest during the first day's lecture and as a posttest during lecture in the last week of class. The Lawson test was given as a posttest at the same time as the CLASS posttest. The populations of each section were assumed to have similar science reasoning pretest knowledge, and it was assumed the semesters represented a random sample of students. Students in the experimental semesters completed additional 30-minute video-based activities as part of their lab sessions. This "instructional modification" consisted of eight short videos designed to improve students' scientific reasoning skills and attitudes toward science. Videos were chosen as the pedagogy because the authors believed video activities would get more engagement from students.

The structure of all the videos was similar and each focused on a different physics topic that was relevant to the course material for that week. The basic structure of the videos was as follows: an introduction to that week's topic; discussion of relevant variables that play a role in that physical system; a mathematical derivation and/or experimental demonstration that test these variables; a summary of the findings on how the variables discussed affect the physical system; a fun clip from YouTube that illustrated the weekly topic with higher production value video. These videos are described in detail in our previous work [24].

Every video was accompanied by a worksheet that students completed and then submitted for a completion grade. There were five questions per video, on average. The worksheet questions varied in style (e.g. multiple choice, free response, or calculation) depending on their purpose. All videos focus on discovering what variables affect a property of the system, control of variables while experimenting, and the correct reasoning from an observation.

III. FINDINGS

A. Effects of the Intervention

Initial results suggested improvement in both scientific reasoning skills and students' attitudes towards science in the first experimental semester compared to the first control semester. A statistically significant difference of 4.7% occurred in the mean Lawson posttest score between the control and experiment semesters. A reduction of the drop in the pre-post CLASS score indicated a 5% improvement in student's attitudes compared to the control semester. Continued use of the intervention over multiple semesters showed sustained higher posttest Lawson scores for courses that used the new curriculum compared to courses that did not. However, the pre-post improvement of students' attitudes that was initially observed was not sustained across all experimental semesters.

Overall, the average posttest Lawson reasoning score was 55% for the two control semesters and 61% for the three semesters that used the video intervention. This is a statis-

TABLE III. Posttest CLASS Expert-like Scores by Attitude Category with (Change Pre to Post - Positive is pre-post shift toward expert.)

Category	# of Items	Control 1	Control 2	Avg. Control	Exp.1	Exp. 2	Exp. 3	Avg. Exp.
Personal Interest	(6)	32% (-13%)	24% (-9%)	30% (-12%)	39% (-7%)	35% (-9%)	50% (+2%)	39% (-6%)
Real World Connection	(4)	41% (-16%)	33% (-10%)	39% (-14%)	43% (-11%)	22% (-31%)	46% (-7%)	36% (-17%)
Problem Solving General	(8)	45% (-10%)	48% (-2%)	46% (-7%)	52% (-6%)	41% (-13%)	59% (+6%)	49% (-6%)
Problem Solving Confidence	(4)	51% (-10%)	49% (-2%)	50% (-7%)	58% (-9%)	37% (-16%)	63% (+6%)	51% (-9%)
Problem Solving Sophistication	(6)	30% (-2%)	54% (+20%)	37% (+5%)	32% (-9%)	34% (-54%)	44% (+7%)	35% (-9%)
Sense Making/Effort	(7)	45% (-17%)	45% (-7%)	45% (-14%)	55% (-5%)	39% (-46%)	49% (+3%)	48% (-18%)
Conceptual Understanding	(6)	29% (-10%)	39% (+9%)	32% (-4%)	40% (-6%)	28% (-12%)	39% (+16%)	35% (-4%)
Applied Conceptual Understand.	(7)	24% (-3%)	55% (+20%)	33% (+4%)	29% (-6%)	20% (-53%)	50% (+11%)	29% (-19%)
All Items (Agree/Disagree)	(36)	38% (-10%)	44% (+2%)	40% (-6%)	44% (-5%)	34% (-27%)	50% (+6%)	41% (-10%)

tically significant difference to students' reasoning scores of 6% (p = 0.011, d = 0.33). In addition, all three of the experimental semesters individually outperformed the two control semesters. This suggests that the intervention curriculum was successful at improving student scientific reasoning skills assuming that the populations had equivalent scientific reasoning skills at the start of each semester.

On the other hand, there was no overall improvement to expert-like attitude scores for students using the experimental curriculum vs. the control curriculum. For both the CLASS posttest (Post-CLASS) and the change from pretest to posttest (Change CLASS), students using the experimental curriculum did not have consistent, significant improvement in their expert-like attitudes. The mean posttest expert-like score was 40% for the Control group vs. 41% for the Experiment group; these were not significantly different (p = 0.42). The mean pre-post change was -6% for the Control group vs. -10% for the Experimental group and again not significantly different (p=0.35) (see Table II). CLASS scores were also investigated by category, but no meaningful or interesting trends were observed (see Table III). (The CLASS was scored by the percentage of expert-like response on the 36 agree/disagree questions. Student who agreed on a question an expert would agree with were given a 1 and likewise for a disagree question. If they chose neutral or opposite to expert-like they were given a zero. This is the "percent favorable" scoring mechanism used most in the literature [2, 15, 16].)

One notable finding from our study was the inconsistency of these students' CLASS scores on a pre- and post- level. There was extensive variation in both pre- and post- mean expert-like CLASS scores between semesters. Pre-CLASS expert-like attitude scores varied from 42% up to 61%, post expert-like scores from 34% to 50%, and pre-post shifts varied from -10% to +2% for the Control semesters and from -27% to +6% for the Experiment semesters (see Table II and Table III). Prior research with the CLASS has not shown such large differences within a specific course and curriculum and generally has indicated that student's attitudes were fairly consistent with time [2, 8, 13–16, 19, 20]. For example, in the study "Correlating student beliefs with student learning using the Colorado Learning Attitudes About Science Survey"

[8], the CLASS was used to measure student beliefs at the beginning and end of several introductory physics courses. This paper showed results from 6 different courses with pre-post shifts ranging from -10% to +2% [8]. In another study, Physics by Inquiry was assessed at 5 different institutions for its effect on student attitude. Pre-post shifts on the CLASS ranged from 2% to 25% [19]. Each of the five studies given in citations 8 - 12 all indicated relatively stable scores despite all reporting on multiple semesters of data to either test curriculum or course intervention at multiple institutions, with multiple instructors, or to investigate the differences in student attitudes from different kinds of institutions, student populations, and course types [8, 13, 14, 19, 20]. In addition, a meta-analysis by Madsen, McKagan, and Sayre, showed gains across a wide range of studies that ranged from -10% to +17% for different kinds of interventions and populations. However, it is notable that only one of these studies, including all those sited in Madsen et. al's paper, was an algebra-based, semi-traditional course, for a general education population, with non-modeling curriculum. Thus, without a larger intersectional cohort to compare to, it is unclear whether the variation we observed with our student population is indicative of general education physics students, our specific institution's students, or caused by another factor.

B. Correlations between Reasoning & Attitude

In considering the effect of the experimental curriculum on students' science reasoning skills and attitudes, we asked the question: Are expert-like attitudes correlated with scientific reasoning? To assess this, correlation coefficients were calculated for each of the five semesters' post-CLASS scores and the Lawson scores (see Table IV). As the scatter plot in Figure 1 clearly indicates, there were no significant correlations between the two posttests for either the control or experiment conditions. Table IV reports the correlation coefficient values. (All correlations greater than ±0.3 are indicated with a *.) The control semesters' correlations were 0.17 and 0.02. Two of the experiment semesters had small positive correlations between CLASS scores and Lawson scores (0.36 and

TABLE IV. Correlations of Post-CLASS Expert-like Score by Category with Lawson Posttest Score. * Indicates correlation greater than 0.3.

Category	# of Items	Control 1	Control 2	Overall Control	Exp.1	Exp. 2	Exp. 3	Overall Exp.
Personal Interest	(6)	0.116	-0.044	0.058	0.139	-0.043	-0.324*	-0.034
Real World Connection	(4)	0.060	-0.135	-0.005	0.162	0.023	-0.175	0.013
Problem Solving General	(8)	0.101	0.060	0.092	0.254	0.105	-0.351*	0.075
Problem Solving Confidence	(4)	-0.029	0.050	-0.007	0.276	0.211	-0.094	0.136
Problem Solving Sophistication	(6)	0.187	0.139	0.189	0.409*	0.163	-0.212	0.207
Sense Making/Effort	(7)	0.136	-0.125	0.059	0.181	0.266	-0.540*	0.014
Conceptual Understanding	(6)	0.151	-0.206	0.065	0.324*	0.179	-0.406*	0.106
Applied Conceptual Understand.	(7)	0.187	0.157	0.191	0.525*	0.161	-0.164	0.206
All Items (Agree/Disagree)	(36)	0.166	0.020	0.139	0.361*	0.300*	-0.571*	0.127

0.30), but the other experiment semester had a negative correlation of -0.57. When taken for all student's in the experiment condition, the correlation was small and almost identical to the control average at 0.13. This data suggests that there is not a strong connection between expert/novice attitudes and scientific or formal operant reasoning for this population of students. This is surprising as one would tend to believe that expert-like thinking would be accompanied by above average ability in scientific reasoning.
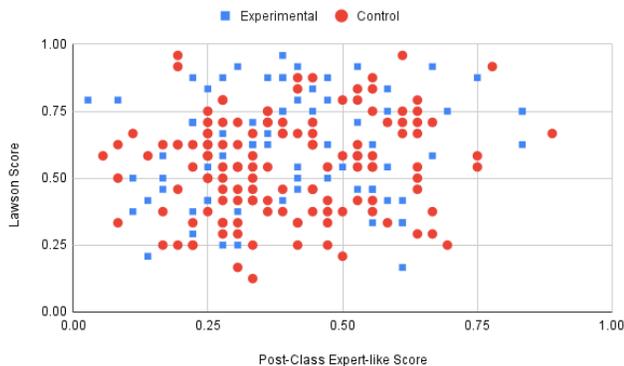

FIG. 1. Posttest Lawson Score vs. Post-CLASS Expert-like Score. This figure shows the lack of correlation between these measures.

A limitation of this data is that because pretest Lawson scores were not collected, student attitude could not be correlated with pre-post gains in Lawson score. This along with expanding the sample to other course levels, such as calculus-based, are important future investigations.

The fourth and final question this research addresses is: Are certain categories of attitudes, such as "problem solving sophistication" or "personal interest", more highly correlated with reasoning expertise? Of the eight different reasoning categories on the CLASS the two categories that correlated most highly with Lawson scores were "problem solving sophistication" and "applied conceptual understanding". This was the case for both control and experiment groups, but

these correlations were not strong. "Problem solving sophistication" correlation coefficients were 0.19 for control and 0.21 for experiment and "Applied Conceptual Understanding" were 0.16 for control and 0.21 for experiment.

IV. CONCLUSIONS

Here, data are presented that indicates that for the student population studied the correlation between students' skill at scientific reasoning, as measured by the Lawson test, and their attitudes toward science, as measured by the CLASS are at best weakly correlated. Not only were low or no correlation between the two measures found, there were also indications that it was possible to improve reasoning skills through a curricular intervention without improving students' attitudes toward science. This suggests that expert-like attitudes are not necessary for the development of reasoning. These findings are consistent with findings from a 2015 article that showed that attitude improvements were not necessarily associated with research based instructional reforms. Inclusion of modeling or other pedagogical course changes were usually necessary to see positive gains in expert-like attitudes [9].

From a teaching perspective, this work suggests direct assessment of student attitudes is important if development of scientific attitudes is an instructional goal. If the Lawson test alone had been used to assess this curriculum the results suggest that the intervention was a successful laboratory addition. However, the lack of expert-like attitude improvements, especially in a general education course, dampens enthusiasm for the curriculum. The authors recommend that more work needs to be done to investigate the mechanisms for improving students' expert-like attitudes. Tremendous progress has been made in PER about the best ways to address conceptual, problem solving, and reasoning instruction. However, relatively little fundamental research is available for how to positively affect students' expert-like attitudes.

This research also speaks to the importance of longitudinal assessment and the dependability of results from the CLASS. The variability of student attitudes in our study indicates that attitudes may need to be measured over multiple semesters to validate a curriculum.

[1] A. E. Lawson, The development and validation of a classroom test of formal reasoning. Journal of Research in Teaching Science 15 (1), 11-24 (1978).

[2] W. Adams, K. Perkins, N. Podolefsky et al., New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. Phys. Rev. ST Phys. Educ. Res., 2, 101010 (2006).

[3] J. C. Moore and L. J. Rubbo, Scientific reasoning abilities of nonscience majors in physics-based courses, Phys. Rev. ST Phys. Educ. Res. 8, 010106 (2012).

[4] J. Hawkins and R. D. Pea, Tools for bridging the cultures of everyday and scientific thinking, J. Res. Sci. Teach. 24, 291 (1987).

[5] L. Bao, et. al., Learning and scientific reasoning, Science 323, 586 (2009).

[6] C. Zimmerman, The development of scientific thinking skills in elementary and middle school, Dev. Rev. 27, 172 (2007).

[7] M. Lovelace and P. Brickman, Best Practices for Measuring Students' Attitudes toward Learning Science. Life Sciences Education, 12, 606-617 (2013).

[8] K. K. Perkins, W. K. Adams, S. J. Pollock, N. D. Finkelstein, and C. E. Wieman, Correlating student beliefs with student learning using the colorado learning attitudes about science survey, AIP Conf. Proc. 790, 61 (2005).

[9] A. Madsen, S. B. McKagan & E. C. Sayre, How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies. Phys. Rev. ST Phys. Educ. Res., 11(1), 010115, (2015).

[10] E. F. Redish, J. M. Saul, & R. N. Steinberg, Student expectations in introductory physics. American journal of physics, 66(3), 212-224 (1998). https://doi.org/10.1119/1.18847

[11] I. A. Halloun, in Proceedings of the International Conference on Undergraduate Physics Education, College Park, Maryland, 1996 unpublished, https://doi.org/10.1063/1.53156.

[12] Elby, A. (2001). Helping physics students learn how to learn. American Journal of Physics, 69(S1), S54-S64, https://doi.org/10.1119/1.1377283.

[13] V. Otero and K. Gray, Attitudinal gains across multiple universities using the physics and everyday thinking curriculum, Phys. Rev. ST Phys. Educ. Res. 4, 020104 (2008).

[14] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Reforming a large lecture modern physics course for engineering majors using a PER-based design, AIP Conf. Proc. 883, 34 (2007).

[15] K. A. Slaughter, S. P. Bates, and R. K. Galloway, A longitudinal study of the development of attitudes and beliefs towards physics, PERC Proceedings AIP Conf. Proc. 1413, 359 (2012).

[16] W. K. Adams, K. K. Perkins, M. Dubson, N. D. Finkelstein, and C. E. Wieman, The Design and Validation of the Colorado Learning Attitudes about Science Survey, PERC Proceedings AIP Conf. Proc (2004).

[17] J. De la Garza and H. Alarcon, Assessing Students' Attitudes In A College Physics Course In Mexico. Physics Education Research Conference (2010).

[18] V. Sawtelle, E. Brewe, L. Kramer, Validation study of the Colorado Learning Attitudes about Science Survey at a Hispanic-serving institution. Physical Review Special Topics-Physics Education Research 5, 023101 (2009).

[19] B. A. Lindsey, L. Hsu, H. Sadaghiani, J. W. Taylor, and K. Cummings, Positive attitudinal shifts with the physics by inquiry curriculum across multiple implementations, Phys. Rev. ST Phys. Educ. Res. 8,010102 (2012).

[20] E. Brewe, A. Traxler, J. de la Garza, and L. H. Kramer, Extending positive CLASS results across multiple instructors and multiple classes of modeling instruction, Phys. Rev. ST Phys. Educ. Res. 9, 020116 (2013).

[21] K. A. Douglas, M. S. Yale, D. E. Bennett, M. P. Haugan, and L. A. Bryan, Evaluation of Colorado Learning Attitudes about Science Survey, 10, 020128 (2014).

[22] 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).

[23] C. C. Tsai and C. Chou, Diagnosing students' alternative conceptions in science, J. Comput. Assist. Learn. 18, 157 (2002).

[24] A. Sammons, J. Tolmie, R. Rosenblatt, and R. Zich, Effect of supplementary videos on scientific reasoning in a general physics course, PERC Proceedings AIP Conf. Proc. (2017).