Exploratory analysis of students' open-ended responses describing their perception of course inclusivity in an introductory physics course

Vanessa P. Bustamante, Hannah E. Blomgren, Dasha K. Walker,
Joshua D. Edwards,[*] and Regina F. Frey[†]
Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, USA

(Received 5 March 2024; accepted 18 July 2024; published 26 August 2024)

In this exploratory study, we examined students' perceptions of inclusion in a calculus-based, introductory physics 1 course for science and engineering majors. This course, offered at a large R1 institution in the United States, was conducted remotely due to the COVID-19 pandemic. Via a survey given at the end of the semester, students rated their course inclusion and provided open-text explanations for their ratings. On average, students rated the course as moderately inclusive. Using inductive qualitative content analysis, six categories emerged: academic, identity, lack of understanding, nonspecific, other, and remote learning. The top three categories were academic (41%), nonspecific (33%), and remote learning (18%). The remote learning category included phrases containing remote learning, Zoom, or COVID-19, along with a second idea explaining the student's level of inclusion, leading to remote learning subcategories. These subcategories were similar to the other primary categories and the academic subcategories. Many students cited academic reasons for their inclusivity scores, including course structure, teaching practices, instructor-student interaction, student-student interaction, and overall course environment. Importantly, many of these factors are within the instructor's influence. Chi-square tests indicated that students perceiving high inclusion emphasized academic factors, while those feeling low inclusion focused on the remote learning aspect of the course. Overall, our findings suggest that instructors can significantly influence students' perceptions of inclusion through various teaching strategies, interactions between instructors and peers, and a welcoming environment. These insights contribute to the ongoing discussion about creating inclusive classrooms by incorporating student perspectives.

DOI: 10.1103/PhysRevPhysEducRes.20.020112

I. INTRODUCTION

As interest in diversity, equity, and inclusion (DEI) in science, technology, engineering, and mathematics (STEM) classrooms increases nationwide, there has been a surge in studies exploring its impact on social aspects of learning and student performance. Past research has shown that motivational factors (e.g., social belonging, self-efficacy, and academic mindset) influence course performance, interest, and persistence in STEM [1–5]. These factors particularly affect marginalized groups (e.g., racial or ethnic groups and women) in physics, as well as other STEM courses [6]. Societal stereotypes also play a crucial role, as women frequently exhibit significantly lower physics self-efficacy and lower physics belonging than men, even when their performance is comparable [2,7]. This can lead many women to avoid pursuing physics-related majors and careers [4,8,9]. However, the recognition by instructors of a student's potential for success in a physics course has proven to be valuable, benefiting all students. Instructors' recognition holds particular significance for underrepresented groups, including women and ethnic and racial minorities, partially due to societal stereotypes in physics and the absence of sufficient role models [10–14]. Therefore, creating equitable and inclusive courses involves considering multiple facets shown to enhance students' performance and retention in physics and STEM in general. The current study explores students' perceptions of inclusion in a large remote-learning calculus-based introductory physics 1 course during the COVID-19 pandemic.

Additionally, in courses like introductory physics 1 (IP1), foundational for science and engineering majors, the broader discipline-based education research emphasizes the interconnected roles of educational, cognitive, and affective factors in shaping students' outcomes in STEM fields. Predictors such as academic history and cognitive measures, including scientific and mathematical reasoning abilities, intertwine with affective measures like student motivation, emphasizing the multifaceted landscape of

[*]Present address: Department of Chemistry, Texas A&M University, College Station, Texas 77843, USA.
[†]Contact author: gina.frey@utah.edu

introductory STEM education. For example, in physics and chemistry studies, students' subject-specific self-efficacy [15,16] and course-level social belonging [2,17–19] have been shown to impact their course outcomes (i.e., performance and retention), including in introductory physics. Additionally, several studies have shown that some of these motivational factors (e.g., social belonging, self-efficacy, and academic mindset) are affected by the classroom environment [15,20]. This complexity suggests that multiple approaches are needed to support students on their academic journeys.

Recent reviews have outlined evidence-based inclusive strategies in STEM, offering suggested implementations [21–23]. These strategies, supported by studies investigating their impact on student performance and persistence in STEM, have contributed to the development of frameworks such as the one proposed by White *et al.* In this framework, inclusive teaching practices include active learning, group work, fostering a sense of belonging, validating students' scientific identities, promoting a growth-mindset culture, proactively reaching out to students, and building relationships [23]. Moreover, a recent policy forum in *Science*, by Handelsman *et al.*, advocates for "fixing the classrooms" and offers actionable steps for instructors to create more inclusive learning environments [24]. While the literature supports a connection between these teaching practices and improved student outcomes, Dewsbury and Brame underscore the importance of involving student voices in discussions about designing inclusive classrooms [21].

Studies by Singh, Cwik, and Li, conducted in calculus-based introductory physics courses, have defined the inclusiveness of the learning environment as comprising belonging (sense of belonging), perceived recognition, and peer interaction [25–27]. Sense of belonging, as defined by Walton and Cohen [28], involves the perception of acceptance, value, inclusion, and encouragement by others. Various studies have highlighted how students' sense of belonging is influenced by their interactions with peers, crucial elements contributing to the inclusiveness of the learning environment [7,29–31]. Additionally, the role of perceived recognition, particularly in the context of students identifying as "physics persons," emerges as a significant factor in shaping the inclusive classroom [7,29]. Furthermore, other studies have noted the importance of student-instructor interaction and its role in creating an inclusive classroom [23,32–34]. Moreover, Hoehn and Finkelstein's exploratory study ($N = 120$) involving a predominantly male (75%) and White (67%) sophomore-level engineering classroom described, via unprompted responses to feedback surveys, students' perceptions and experiences of inclusive teaching practices. These responses included group collaboration, course structure, having people they know in the course, and having teachers listen to student feedback and implement change [20], which align with the inclusive classroom constructs found in Singh, Cwik, and Li's work.

While IP1 at this institution is traditionally taught in person, the onset of the Covid-19 pandemic forced this course, as with many courses, into an online setting, requiring students and instructors to quickly modify course structure and course communication approaches. This quick change to online introduced a new challenge: replicating physical classrooms online using preexisting systems and resources that institutions already had in place for delivering education in remote or online formats [35]. In addition to the type of online delivery method used, effective teaching practices and institutional administrative support are required to sustain meaningful student engagement [36]. For example, online course communication should involve three types of interactions: student-student, student-content, and student-instructor [37]. While various studies suggest different pedagogical methods of online course delivery (e.g., the technological pedagogical content knowledge model [38,39] or the C♭-model [40]), there is little research on what students perceive as comprising an inclusive online environment. For example, Petersen conducted a literature review of best practices for an inclusive online learning environment and recommends teaching students time management and using calendars to keep track of important course deadlines [41]. Additionally, the review recommends that instructors be more intentional with structuring their learning environment, course activities, and curriculum [41].

II. RESEARCH QUESTIONS

Building upon Dewsbury and Brame's call for student involvement, this research contributes to the ongoing conversation by soliciting students' insights and exploring the characteristics and teaching practices identified by these voices that contribute to their perception of inclusivity in an introductory physics 1 course. In this study, we explored written student responses to the Likert-scale question, "How inclusive do you feel the course [course name] was overall?" and the following open-ended question, "Please explain your response to the previous question." Employing inductive qualitative content analysis on the student responses, we formulated three research questions (RQs):

RQ1. What primary themes emerged from students' descriptions of the inclusivity in their first-semester introductory physics course? Additionally, what themes (within the primary categories) are most reported by different identity groups (gender, race or ethnicity, and first generation), those who report low/high course inclusion, and final-course grades?

RQ2. *Post hoc* question: What academic subthemes emerged from students' descriptions of the inclusivity in their first-semester introductory physics course? Additionally, what subthemes (within the Academic subcategories) are most reported by various identity groups (gender, race or ethnicity, and first generation), those who report low/high course inclusion, and final-course grades?

RQ3. Given that this study was conducted during the pandemic, what themes emerged from students' descriptions of how being in a remote-learning course for their first-semester introductory physics course affected their perception of inclusivity? Additionally, what subthemes (within the remote learning subcategories) are most reported by different identity groups (gender, race or ethnicity, and first generation), those who report low/high course inclusion, and final-course grades?

III. METHODS

A. Study setting

This study took place in Fall 2020 at a large Mountain West, research-intensive university in the United States during the Covid-19 pandemic. In the Fall 2020 semester, approximately 25 000 undergraduate students were enrolled at this university.

The study was conducted in a first-semester introductory physics course (IP1), which is the first part of a two-course calculus-based sequence in introductory physics required for many physical science and engineering majors. On average, IP1 typically enrolls between 400 and 600 students during the fall or spring semester of the academic year. Course topics involved classical mechanics, such as kinematics, dynamics, energy, momentum, rigid body rotation, oscillations, and waves. Because of the Covid-19 pandemic, in Fall 2020, this course was offered as an exclusively synchronous online course (i.e., no in-person meetings), which included two 80-min lectures and two 50-min discussion sessions each week. The discussion sections, in which students worked collaboratively in small groups to solve problems via breakout rooms, were led by graduate teaching assistants (TAs) with the support of undergraduate learning assistants (LAs). LAs at this institution are paid undergraduate students who are formally trained through a semester-long pedagogy course and work with small groups of students to facilitate collaborative active learning during the course based on the nationally recognized LA model [42].

The course consisted of two sections with each section taught by different professors, one woman and one man. In section A, discussion attendance was optional, and in section B, attendance at discussion sessions was required to receive participation points. Both sections required students to watch prelecture videos and complete preassignments before each lecture session. Students in section A were also required to complete postlecture assignments.

As shown in Table I, section A students were evaluated based on pre-and postlecture assignments (10% total grade), weekly homework assignments (11 in total, contributing 25% to the overall grade), quiz scores (11 quizzes in total, with the lowest score dropped, accounting for 10% of the total grade), midterm exams (2 in total, with 20% for each exam, making up 40% of the total grade), and a final project (15% of the total grade). The final project entailed the student choosing and describing five examples of physics topics covered in the course and applying each to real-life scenarios [2].

Section B student grades (see Table I) were determined by the following components: discussion sections (10% of total grade), preclass assignments (26 in total, with the lowest 6 dropped, 17% of total grade), weekly homework assignments (12 in total, with the lowest 2 dropped, 20% of total grade), quiz scores (12 in total, with the lowest 3 dropped, 10% of total grade), midterm exams (2 in total, with 15% assigned to each for 30% of total grade), and a cumulative final exam (20% of the total grade). Each exam, including the final, had two parts: an in-class, online proctored section (90% of the exam grade) and an out-of-class, take-home component (10% of the exam grade). For the midterm exams, the in-class part involved both individual and group work sections.

B. Inclusion questions

During the last 2 weeks of the Fall 2020 semester, an end-of-semester survey was collected online via Qualtrics survey software [43], with demographic questions being asked at the end. The two course-inclusion questions

TABLE I. Section breakdown.

Section A		Section B	
Assignments	% of overall grade	Assignments	% of overall grade
Pre-post lecture assignments	10	Discussion	10
Weekly homework	25	Preclass assignments	17
Quizzes	10	Weekly Homework	20
Midterm exams	40	Quizzes	10
Final project	15	Midterm exams	30
		Final exam	20
Total	100	Total	107

included a five-point, Likert-scale overall course-inclusion question which asked, “How inclusive do you feel the course [insert course name] was overall?” corresponding to the following scale: (1) “Not at all,” (2) “Slightly,” (3) “Somewhat,” (4) “Moderately,” (5) “Highly.” This question was followed by “Please explain your response to the previous question.”

All recruitment and study procedures were approved by the university’s institutional review board. Students received no compensation for study participation but received a small amount of extra credit for completing surveys in the course, independent of their consent to allow their data to be used for research purposes.

C. Participants

A total of 554 students enrolled in IP1 during the Fall 2020 semester. The IP1 sections consisted of approximately 17% first-year students, 33% sophomores, and 50% upper-level students (i.e., juniors, seniors, and postbaccalaureate students). Of these students, 404 provided consent, and from this sample, 366 students provided a Likert-scale rating to the course-inclusion question. Of the 366 students, 289 students (79%) provided both a response to the Likert-scale inclusion question and a response to the open-ended portion of the question. Of the 77 students (21%) who did not provide a response, 73 left the question blank and 4 wrote NA or “no comment.” Because we wanted to examine the reasons for the student’s perceptions of inclusivity, we only analyzed the responses from the 289 students. However, we did compare the average Likert-scale ratings for the students who did and did not provide an open response and found them to be comparable ($M = 4.41$, $M = 4.22$, respectively, which is a “moderately inclusive” rating; $t_{112} = -1.88$, $p = 0.063$). See Supplemental Material Sec. A [44] for detailed descriptive statistics for the course-inclusivity rating for both sets of students.

D. Demographics and course grades

Demographic information for students was collected through their survey responses, focusing on three variables: *gender, first-generation status, and race or ethnicity* (see Table II). In response to the question, “To which gender identity do you most identify?” students selected from options: “Man” ($n = 202$), “non-binary/third gender” ($n = 0$), “Woman” ($n = 86$), “prefer not to answer” ($n = 1$), or “prefer to self-describe” ($n = 0$). To determine first-generation status, students were asked, “Did any of your parents/legal guardians obtain a college degree?” with responses including “Yes” ($n = 200$), “No” ($n = 66$), or “I do not know/prefer not to answer” ($n = 23$). In response to the question, “I identify my race/ethnicity as: [*Select all that apply.*],” students indicated their identity with options: “Asian” ($n = 37$), “Black/African American” ($n = 1$), “Hispanic/Latinx” ($n = 19$), “Multi-racial” ($n = 20$), “Native American/Alaskan Native” ($n = 1$), “Pacific Islander” ($n = 1$), “White” ($n = 194$), “Prefer not to answer” ($n = 12$), or “Prefer to self-describe” ($n = 4$). We recognize the limitations of the term “Asian,” as students under this category may come from diverse cultural and ethnic backgrounds, leading to substantially different experiences in physics and STEM in general [45,46]. The authors also recognize the importance of representing the many varied experiences and individual, distinct identities of persons of color in STEM. However, the marginalized categories of race or ethnicities were gathered into one variable to aid in building a baseline to better understand their course inclusivity in physics [6,47,48]. For this marginalized population of students, we are using the term racially marginalized (RM), which consists of the student population who identify as Black/African American, Hispanic/Latinx, multiracial, native American/Alaskan native, or Pacific Islander. Four students preferred to “self-describe” their race or ethnicity as the

TABLE II. Overview of student demographics.

	Student demographics							
	First-generation status			Gender				
	non-FG	FG	Prefer not to answer/I do not know	Men	Women	Nonbinary	Prefer not to answer	Prefer to self-describe
n	200	66	23	202	86	0	1	0
%	69%	23%	8%	70%	30%	0%	0%	0%

	Race or ethnicity							
		Racially marginalized (RM) in STEM ($n = 45$, 11.7%)						
	White	Black/African American	Hispanic/Latiné	Multiracial	Native American/ Alaskan Native	Pacific Islander	Asian	Prefer not to answer
n	194	1	19	20	1	1	37	13
%	70.2%	>1%	6.9%	7.2%	>1%	>1%	13.4%	4.7%

following identifiers: Biracial, Human, Latino, Biracial, and Middle Eastern (Arabic). The student identifying as "Human" was placed into "Prefer not to answer" because they selected all race or ethnicity options, while the rest of these students were placed in the "Racially Marginalized (RM)" group, leading to the following sample sizes for the three race or ethnicity categories: Asian ($n = 37$), RM ($n = 45$), White ($n = 194$), and "prefer not to answer" ($n = 13$).

Course grades were obtained from instructors after the final grades were reported, for students who gave consent to participate in the research study. Although course grades are not an all-encapsulating way to measure student learning, they provide a gauge of student performance and can have an impact on student perceptions of IP1 [1]. Recent literature has shown that motivational factors, such as belonging and students' perception of their learning environment, play an important role in predicting final course grades for students in an introductory-level algebra-based physics course [7]. In our current study, we analyzed students' final course grades, where group 1, referred to as *GGroup 1* (GG1), consisted of students who earned grades C+ or lower ($< 75\%$). Group 2, known as *GGroup 2* (GG2), comprised students who achieved grades A- or above ($> 89\%$). The "C+ or below" range was selected because a general chemistry 1 study at this same institution found that students whose major required them to take general chemistry 2 and had a C+ or below in general chemistry 1 had less than a 50% probability of continuing on to general chemistry 2 [49].

E. Qualitative analysis

A qualitative methodology was applied to explore the research questions, employing an inductive qualitative content-analysis approach [50,51]. The process, guided by Merriam's principles [52], involves three key steps. First, categories are constructed by inspecting randomized datasets and recording emergent themes relevant to the research question(s). This process continues until a saturation point is reached, indicating the emergence of no new themes. Second, the initial codes undergo iterative sorting, refinement, and grouping into broader categories with evolving definitions. The coding scheme (including definitions) is refined until all study-related themes are sufficiently captured. In the last step, the validity of the codes is affirmed through interrater reliability (IRR), engaging at least two independent raters who code a subset of the responses.

A codebook was developed outlining primary categories and subcategories related to academic and remote-learning aspects, with the goal of understanding how students express their inclusivity within their IP1 course. The structure of the codebook was developed from a similar codebook employed in a simultaneous STEM inclusion study carried out at a private research-intensive institution before the onset of the pandemic [34] and was developed in conjunction with a simultaneous study in general chemistry 1 at this institution [33]. While students were not asked to provide a sole explanation for their Likert-scale rating, their responses typically consisted of a phrase or sentence, each centering on a specific category. Although the coding process allowed for double coding to accommodate two independent ideas, all student responses in our sample discussed a single idea. As a result, all responses were coded into one category, with no instances of double coding. The codebook underwent refinement and expansion, incorporating data from two IP1 courses (Fall 2019 and Spring 2020) and a general chemistry 1 course (Fall 2020) at the same institution. In this study and the general chemistry 1 (fall 2020) study, students also provided inclusivity feedback for remote-learning courses during the pandemic. Therefore, this study helped to enhance the codebook with a remote-learning category and corresponding subcategories. In addition, this study helped to validate the nonremote categories from a different research-intensive institution with a distinct student population using five different introductory STEM courses (biology, chemistry, and physics) and helped to expand the codebook to encompass categories that students identified as relevant in the context of remote-learning courses during the pandemic featuring both synchronous and asynchronous components. See Supplemental Material Sec. B [44] for the complete codebook.

The iterative development process involved two coders, D. W. and H. B., who independently coded randomized sample responses, with the coding results being compared, and any discrepancies in codes were thoroughly discussed before reaching a consensus by the qualitative research team (D. W., H. B., and R. F.). Each team member contributed a unique perspective to the qualitative data analysis: (1) D. W., an undergraduate with a background in psychology and chemistry; (2) H. B., a master's graduate student in teacher education with a physics bachelor's degree and a focus on teaching, including having STEM education-research coursework, and (3) R. F., a Ph.D. in chemistry, a STEM faculty member who teaches general chemistry, and a STEM discipline-based education researcher with expertise in qualitative coding and inclusive teaching within introductory STEM courses. The final categories, definitions, and example quotes are in Tables III–V (see Supplemental Material Sec. B [44] for the complete codebook).

Although both coders, D. W. and H. B., individually coded all responses for IP1 ($n = 289$) and the qualitative research team discussed all discrepancies until a consensus was reached, the IRRs for coded responses were calculated. The IRRs were determined on 30% of the entire dataset using Krippendorff's alpha statistic [53,54], which accounts for the chance of agreement. For the largest of the primary categories, the academic category, the IRR was 0.861 (four discrepancies), and the second largest category

TABLE III. Definitions of the primary categories.

Primary category	Definition
Academic ($N = 118$)	Students referenced teaching practices and course design elements, and the roles they played in how students engaged in and felt a part of the learning process. This was described by discussing the environment, the opportunities students had to participate, and if evidence-based teaching practices were used.
Identity ($N = 10$)	Student referred to a person's identity and the role it played in how involved the students felt in the learning process. This was described through language, course content, or in a capacity related to the individuals involved in the course.
Lack of understanding ($N = 12$)	Students expressed ideas that show a misunderstanding of or not valuing the incorporation of inclusion and equity in a STEM course.
Nonspecific ($N = 94$)	Students described their experience of inclusion in the course but gave no specific details.
Other ($N = 4$)	Student's comment did not relate to inclusion in the classroom.
Remote learning ($N = 51$)	Student referenced Zoom lecturing, online or hybrid learning, synchronous or asynchronous classes, Covid-19, or remote learning.

TABLE IV. Definitions of the academic subcategories.

Academic subcategory	Definition
Course structure ($N = 39$)	Students referenced the aspects of the course design or infrastructure that affected the inclusivity.
Environment ($N = 49$)	Students noted the classroom environment with respect to how comfortable the class felt, the amount of collaboration with other students during lecture, the amount of respect there was in class, how engaging the lecture was, how supported the students felt, or the overall atmosphere of the classroom.
Student-instructor interaction ($N = 30$)	Students addressed the ways in which the instructor(s) encouraged participation, how equal they felt participation was, or how comfortable they personally felt answering/asking questions in the classroom.

TABLE V. Definitions of the remote-learning subcategories.

Remote-learning subcategory	Definition
RL-course structure ($N = 17$)	Students referenced the aspects of the course design, course structure, or teaching practices that affected inclusivity in an online or hybrid course.
RL-environment ($N = 9$)	Students noted the course environment, in the context of remote or hybrid learning, with respect to how comfortable the course felt, the amount of respect there was in the course, how engaged the students were, how supported the students felt, or the overall atmosphere of the course.
RL-nonspecific ($N = 20$)	Students gave a nonspecific response that referenced online learning.
RL-student-instructor interaction ($N = 2$)	Students addressed the ways in which the instructor(s) encouraged participation, how equal they felt participation was, or how comfortable they personally felt answering or asking questions in a synchronous classroom setting.
RL-student-student interaction ($N = 3$)	Students noted the nature of their interactions with their peers in an online or hybrid setting.

was nonspecific, with an IRR of 0.753 (two discrepancies). While data were collected for IP1, a similar study was conducted on another STEM course (general chemistry 1) within the same research university. As such, developing the academic subcategories for IP1 was done simultaneously with the general chemistry 1 data, which were 0.70 and above. When calculating the IRRs for the remote learning subcategories, we found that all alphas fell below

0.80 but above 0.60. The threshold level is 0.67 for exploratory studies [55]. However, the coders, D.W and H.B, coded all the responses individually and the qualitative research team agreed on the coding of all responses. See Supplemental Material Sec. C [44] for detailed information on the IRRs.

F. Quantitative frequency analysis

Once all the responses ($n = 289$) were coded into one of the six emergent categories (academic, identity, lack of understanding, nonspecific, other, and remote learning) and subsequent subcategories, we determined the frequency of responses in each category and subcategory and compared these frequencies by gender, race or ethnicity, first-generation status, students reporting low or high course inclusion (low inclusion: score 1 [Not at all]–3 [somewhat]; high inclusion: score of 5 [highly]), and course grade (GGroup 1: C+ or below; GGroup 2: A- or higher).

To establish a baseline understanding of factors identified by students reporting different levels of course inclusivity, we examined the responses of students reporting inclusion levels at the two extremes of the inclusion scale. We excluded, from the frequency analyses, students who gave an inclusion score near the class average (Mean $= 4.40$, SD $= 0.79$), which were those students who rated IP1 as (4) moderately inclusive ($n = 85$). This exclusion of the "middle" group ensured a clear distinction between the two inclusion-reporting groups. Therefore, students reporting a score of 1 [not at all]–3 [somewhat] were placed into the "low" inclusion group, while students reporting a score of 5 [highly] were placed in the "high" inclusion group. See Supplemental Material Sec. A [44] for descriptive statistics for course-inclusivity rating, including a histogram of student ratings and percent and number of students in each rating score.

To determine any variations across the subgroups, we conducted a chi-square test of independence (or Fisher's exact test for sparse data) with an alpha level of 0.05. The chi-square test of independence is used to compare the distribution of the categorical variables between two or more groups through an approximation [56]. If there is no difference between the groups, this suggests that there is no difference in the distribution of the categorical variables for the subgroup of interest. The chi-square statistic is the conventional method for larger sample sizes; therefore, we used a Fisher's exact test with some subgroups for cells with low frequencies or cell counts of zero. To evaluate the effect sizes of the comparisons, we calculated Cramer's V [57] using the guidelines of 0.2-small, 0.5-medium, and 0.8-large [58]. If the chi-square test or Fisher test was significant, then adjusted residuals were calculated to determine each individual cell's contribution to the chi-square value. Next, Bonferroni corrections were computed to determine if the calculated residual was significant to its respective chi-square value, i.e., the cell p values were compared to the Bonferroni's corrected p-value threshold (the original significance-level p value divided by the number of cells) to determine if a subgroup was more likely associated with a specific code. While some criticize the Bonferroni correction for being "overly conservative" [57], many researchers use this correction method to reduce the likelihood of committing a type I error (claiming a false positive). All analyses were run using the base version of Stata Statistical Software: Release 17 [59]. The adjusted residuals and chi-square tests for the residuals were calculated using Microsoft Excel version 16.79.2.

IV. RESULTS

Six primary categories (academic, identity, lack of understanding, nonspecific, remote learning, and other) emerged from the qualitative responses provided by students, and then the responses were further examined for emergent subthemes (i.e., subcategories) for categories containing large percentages of the responses (namely, academic and remote learning). Quantitative results from the survey, in which the students in the study rated the overall level of inclusivity in their course (scale of 1– "not at all" to 5– "highly inclusive") had a high average inclusion score (Mean $= 4.40$, SD $= 0.79$, $N = 289$). Overall, 60% of students rated the course as highly inclusive. As a result, many of the students' responses were written in a positive manner. See Supplemental Material Sec. A [44] for descriptive statistics for course-inclusivity rating, including a histogram of student ratings and percent and number of students in each rating score.

A. Research question 1: Primary themes

Figure 1 displays the distribution of the primary categories (see Table III for definitions and Table VI for example quotes). Overall, we can see that a majority of student responses mentioned academic factors (41%, $n = 118$), followed by nonspecific (33%, $n = 94$), and then remote learning (18%, $n = 51$). The academic category comprises course structure, environment, and student-instructor interactions (Table IV). Students provided a general statement in the nonspecific category without further detail regarding specific factors contributing to their course inclusion (e.g., "idk, its just pretty inclusive."). The category with the third highest response was related to remote learning, in which students described their course inclusivity as being affected by some aspect of remote learning (e.g., Zoom, asynchronous learning, or Covid-19) and then gave a second reason for their level of inclusivity. Details for the remote learning subcategory are discussed later. Factors such as "Identity" (3%), "Lack of understanding" (4%), and "Other" ($< 1\%$) contributed approximately 7% to the overall responses. Responses in the identity category attributed their course inclusivity to factors such as demographics (including disabilities and personality traits), knowledge background, and professor language (e.g., "It works for many but I feel

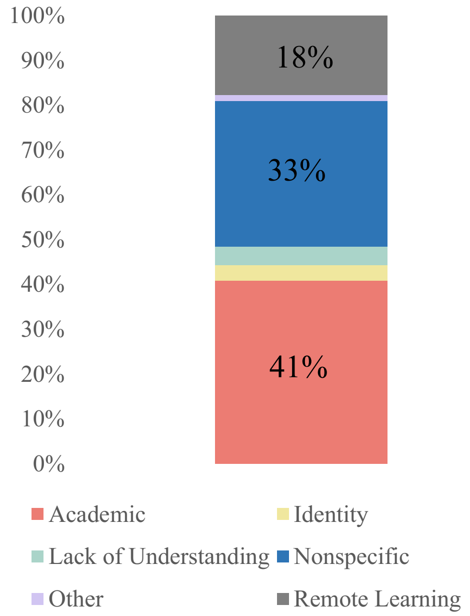

FIG. 1. Response distribution for primary categories. Top three categories are academic (41%, $n = 118$), nonspecific (33%, $n = 94$), and remote learning (18%, $n = 51$).

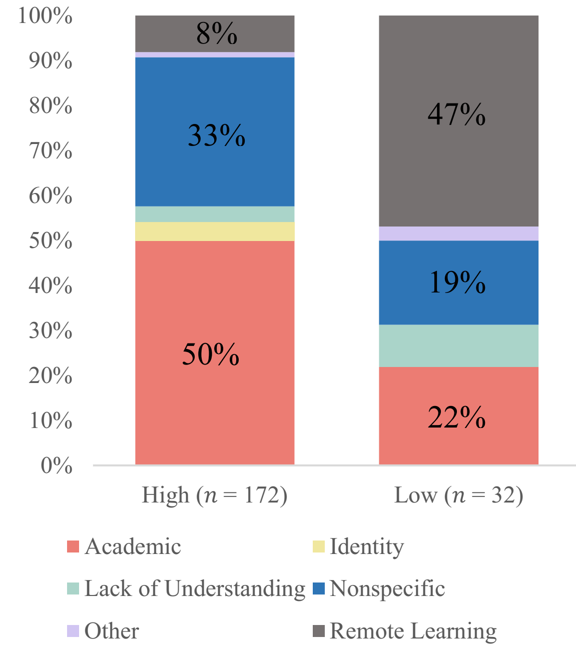

FIG. 2. Distribution of responses for primary categories by reported inclusion score. The top three categories are academic (high: 50%, $n = 86$; low: 22%, $n = 7$), remote learning (high: 8%, $n = 14$; low: 47%, $n = 15$), and nonspecific (high: 33%, $n = 57$; low: 19%, $n = 6$).

like I have preexisting conditions that make this setting more difficult to include myself in."). Students in the "Lack of Understanding" category expressed ideas that showed a misunderstanding of what inclusion in a STEM course meant (e.g., "I think the PHYS 2210 process as a whole covered the basic physical process well."). Finally, students in the Other category provided a comment that did not relate to inclusion in the classroom.

Students possess diverse identities, attitudes, and performance levels that can significantly influence their sense of inclusivity in the classroom. We investigated five subgroups (gender, race/ethnicity, first-generation status, low/high course-inclusivity response, and final course grade) to determine whether certain factors held greater significance for specific groups. Chi-square (or Fisher's exact) tests were conducted on the distributions to assess the significance of differences across these subgroups in the primary categories. Grade groups (GG1 and GG2) exhibited a difference in the distribution [Fisher's exact $p = 0.046$, d.o.f. $= 5$; Cramer's $V = 0.24$ (small effect size)] of the factors they used to describe their course-level inclusion. However, due to a small sample size in GG2 and the uneven sample-size distribution between GG1 ($n = 168$) and GG2 ($n = 20$), we feel it is not reasonable to claim significance. The overall self-reported inclusion score (high or low) also showed a significant trend with a medium effect size (Fisher's exact $p = 0.00000186$, d.o.f. $= 5$; Cramer's $V = 0.44$). The distribution of the high or low inclusion groups in the primary categories is shown in Fig. 2. Adjusted residuals were used to determine the individual cell contributions to the chi-square value, followed by Bonferroni corrections to determine the significance of each cell to the chi-square value. Within the low or high inclusion group, we found that academic and remote learning factors were significant (See Supplemental Materials Sec. D [44] for additional information on adjusted residuals). That is, students reporting low course-level inclusion attributed their reasons to remote learning factors, while students who reported high course-level inclusion, attributed their reasons to academic factors. See Supplemental Material Sec, E [44] for the distributions for all the other subgroup comparisons. See Supplemental Material Sec. F [44] for statistical tests of the nontrending primary categories.

B. Research question 2: Academic subthemes

In IP1, the academic category emerged from 41% ($n = 118$) of the total student responses. To further understand what students were discussing in this factor, the following subcategories emerged from the student responses: course structure, environment, and student-instructor interactions. Tables IV and VI contain definitions and example quotes for the academic subcategories. For students whose responses were coded as academic (Fig. 3), 42% ($n = 49$) were coded as classroom environment, followed by course structure (33%, $n = 39$) and student-instructor interactions (25%, $n = 30$). Similar to the primary categories, we examined the subgroups (i.e., gender, race or ethnicity, first-generation status, low or high course inclusion response,

TABLE VI. Example quotes for the primary categories and academic and remote-learning subcategories.

Primary category	Subcategories	Examples
Academic	Course Structure	"I felt it was highly inclusive because of… growth mindset was taught a lot and mistakes were encouraged, many resources were offered if the students felt like they needed help, there were a lot of options on how to participate in class…"
		"Lots of opportunity to be included and have the chance to be able to speak up and do group work"
	Environment	"I felt very safe and comfortable in the clss [class]. I was able to ask questions and get good feedback!"
		"I never felt like anyone ever did something to exclude someone else. Every effort was made by the instructor and students to make this course as friendly a learning environment as possible."
	Student-Instructor Interactions	"Professor always asked for participation and made everyone feel included even if their ideas were not quite right. This class was designed in a way that made it possible for everyone to succeed if they wanted to."
		"The [class name] teaching team made every effort to be available in study halls for questions. They never made me feel as if I was stupid because I didn't know the answer."
Identity	*Demographics*	"Overall I have found my section of Physics 2210 to be open to all personalities, ages, ethnicities, etc. Professor [P1] consistently affirms that encountering difficulties in learning the material of the course is not unusual which I hope helps people feel less intimidated about taking the course."
	Background knowledge	"While you did need some knowledge in calculus to take this course, it was not so much so that it inhibited everyone from taking this course."
	Professor language	"Everyone got along really well and Prof [P2] was also very good with the use of pronouns and such to ensure that every student felt safe and comfortable"
Lack of understanding	*Attitude*	"I'm not sure what this question is asking. If it's asking if the course includes lots of poc/minority voices, I'd have to say that the question doesn't make sense in the context of a physics course. We did learn about some of the mathematical work the Arabic world did though, so I guess that's something."
	Definition of academic inclusion	"I think the PHYS 2210 process as a whole covered the basic physical process well."
Nonspecific		"Nothing in the course has really seemed exclusive."
Other		"I know I'm doing alright but I really do not enjoy physics and don't plan on taking the next physics class for my math elective in CS."

(Table continued)

TABLE VI. *(Continued)*

Primary category	Subcategories	Examples
Remote learning	RL-course structure	"Professor [P2] was really responsive to how students were feeling and doing during the course. Many times due dates became more flexible and chances to improve test scores were given. I was greatly appreciative of this and I'm sure other students were. Due to the Covid pandemic, myself and other students can be in more difficult life circumstances, and I feel like this course was perfect for people who unable to follow rigid schedule due to things outside of school."
		"It was as inclusive as an online course can get. It definitely wasn't exclusive. We had some group quizzes."
	RL-environment	"It's hard to keep people engaged when they can simply shut off their camera and not talk. But all in all, when people participated it was inclusive."
		"It's probably just a matter of the current situation, but the class definitely felt distant, that's for sure."
	RL-nonspecific	"It is about as inclusive as an online class could ever be"
	RL-student-instructor interactions	"[P1] definitely tries to encourage participation, but I don't really feel like it helps me out when it's over Zoom. In person would def be better"
		"throughout the course, the course staff has been consistent in encouraging participation regardless of perceived academic ability, personality, or other factors. They strove to break down the barriers placed between us by remote learning. I don't feel like the task made the course equal to an in-person experience, but it was much better than many other online courses I have experienced."
	RL-student-student interactions	"Once again online class, not much to really say about inclusivity when there isn't really interactions between people"
		"There were breakout rooms, so you could somewhat get to know people, but just for a few minutes."

and final course grade) to determine if there were factors in the academic subcategories that were described more by some groups. After conducting chi-square tests of independence, we found that no subgroup showed significant differences in their distributions within the academic subcategories. See Supplemental Materials Sec. G [44] for the distributions of responses for subgroup comparisons of academic subcategories whose chi-square tests were not significant. See Supplemental Material Sec. H [44] for statistical tests for nontrending academic subcategories.

C. Research question 3: Remote-learning subthemes

This study took place during the Covid-19 pandemic, and as a result, IP1 was conducted in a hybrid, flipped-classroom learning environment over Zoom. Therefore, we examined what subthemes were present in the composition of students' responses pertaining to a remote learning environment. Of the total responses ($n = 289$), 18% of students' open-ended responses ($n = 51$) discussed their course inclusivity in relation to a remote-learning environment, mentioning words such as "remote-learning," "Zoom," or "Covid-19" in their responses. Tables V and VI contain definitions and example quotes for the remote-learning subcategories.

Within the remote-learning subcategories (i.e., RL-course structure, RL-environment, RL-nonspecific, RL-student-instructor interaction, and RL-student-student interaction), we found that most responses were RL-nonspecific (39%, $n = 20$) (see Fig. 4). A similar trend was seen in a previous study in general chemistry 1 at the same Mountain West University [33]. Most responses in the RL-nonspecific category described that there were efforts made to make the course inclusive. However, due to the online classroom structure enforced by the pandemic, students felt that nothing more could be done to increase inclusivity, thus attributing their main reason for their Likert-scale response to the idea that the course was remote. The following are example responses: "Almost everything was done to make the class

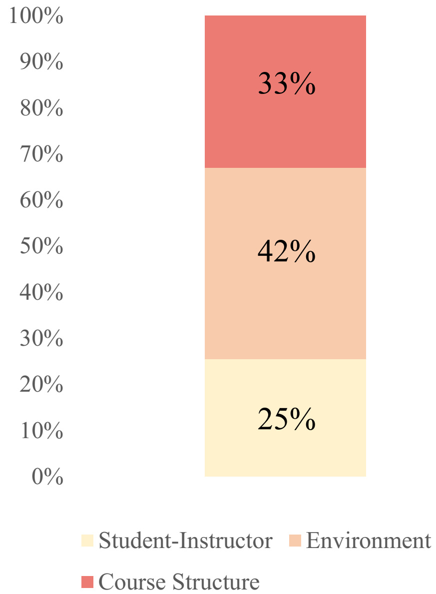

FIG. 3. Response distribution for academic subcategories. Course structure (33%, $n = 39$), environment (42%, $n = 49$), student-instructor interactions (25%, $n = 30$).

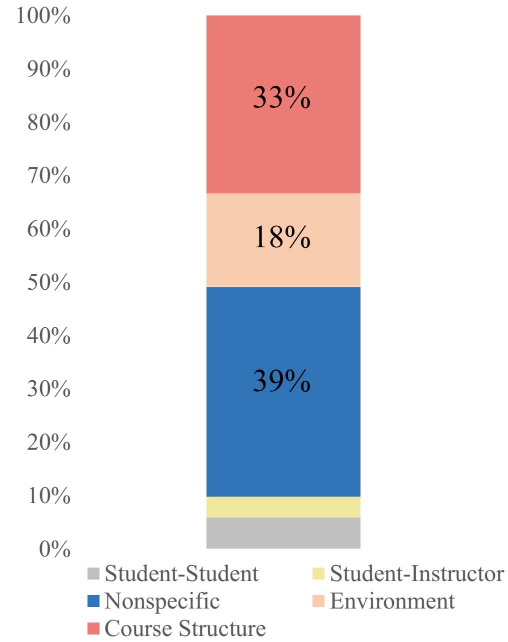

FIG. 4. Distribution of responses in remote-learning category by subcategories. The top three subcategories are RL-nonspecific (39%, $n = 20$), RL-course structure (33%, $n = 17$), and RL-environment (18%, $n = 9$).

more inclusive, the teaching team did a good job, but the online format is really not preferred."; "Because it was an online course, I feel like this dampened the effects."; "I feel like it is hard to be included with the purely Zoom online classes." The next top categories were RL-course structure (33%; $n = 17$), followed by RL-environment (18%; $n = 9$). Example quotes of RL-course structure include "This class has been very inclusive with breakout rooms and many opportunities to ask questions and work with others." and "It is hard to be fully inclusive in online learning and the job falls more on the student." A few example quotes of RL-environment are as follows, "It's probably just a matter of the current situation, but the class definitely felt distant, that's for sure." and "It's hard to keep people engaged when they can simply shut off their camera and not talk. But all in all, when people participated it was inclusive."

Figure 4 shows the percentages of responses in the remote-learning (RL) subcategories for students who reported overall inclusion responses. The RL subcategories were examined across the subgroups (i.e., gender, race or ethnicity, first generation, low or high course inclusivity response, and final course grade) for any variations. See Supplemental Material Sec. I [44] for the distributions of responses for subgroup comparisons of remote-learning subcategories. Gender (men and women) showed differences in their distributions [Fisher's exact $p = 0.037$, d.o.f. $= 4$; Cramer's $V = 0.45$ (medium effect size)]. For example, it seems that men ($n = 33$) and women ($n = 17$) used different factors to describe their course-level inclusion paired with remote-learning factors. Additionally, inclusion (high and low) groups in the remote-learning subcategories showed differences in their distributions [Fisher's exact $p = 0.041$, d.o.f. $= 4$; Cramer's $V = 0.53$ (medium effect size)]. That is, students in the high inclusion group ($n = 36$) seemed to describe different factors than those in the low group ($n = 2$) when discussing their perception of course-level inclusion. However, due to the small sample sizes in the gender and inclusion groups in the remote-learning subcategories, we feel it is not reasonable to claim significance.

V. DISCUSSION AND IMPLICATIONS

The results of this exploratory study in a large calculus-based IP1 course can be summarized in three key findings: (i) the total percentage of students who discussed academic factors (academic primary + RL-environment, RL-course structure, RL-student-instructor interaction) in their responses was 45%. Notably, despite course sections ranging between 150 and 200 students, 97% of all academic-related responses highlighted academic factors within the instructor's sphere of influence; (ii) students who felt that IP1 was highly inclusive (Likert-scale of 5) attributed their reasons to academic factors such as course structure, student-instructor interaction, and student-student interaction, whereas students who felt that their course-inclusivity was low used remote learning factors to explain their reasons; and (iii) Identity comprised only 3% ($n = 10$) of student responses; however, one should not assume that this factor did not influence perception of students' inclusivity. Because the current results reflect a small sample size of $N = 289$, different results could be observed in a course of a larger size. These three findings are discussed below.

While IP1 is typically an in-person class at this institution, students and instructors were forced into a remote environment due to the Covid-19 pandemic during the Fall 2020 semester. As a result, course content and instructional methods were quickly changed, introducing new obstacles for instructors and students. In this exploratory study, student open-ended responses ($N = 289$) were analyzed to probe what factors students used to describe their perception of course inclusivity. Based on students' Likert-scale (five-point) ratings of how inclusive they felt IP1 was, our findings indicate that over half of the students found IP1 to be highly inclusive (60%). Using inductive, qualitative content analysis, six categories emerged from the student responses: academic, identity, lack of understanding, nonspecific, remote learning, and other. From the largest category, the academic category, three subcategories emerged: course structure, environment, and student-instructor interactions. Additionally, six remote-learning subcategories emerged from a more in-depth examination of the effect a remote-learning environment had on students in IP1: Four subcategories (RL-environment, RL-course structure, RL-student-instructor interaction, and RL-student-student interaction) describe factors similar to the academic subcategories, with the other two subcategories being RL-nonspecific and RL-lack of understanding; there was no RL-identity subcategory.

A. Many students attributed their course inclusivity to academic factors

The total percentage of students who discussed academic factors (academic primary + RL-environment, RL-course structure, RL-student-instructor interaction, and RL-student-student interaction) in their responses was 52%. Notably, despite course sections ranging between 150 and 200 students, 97% of responses specifically highlighted academic factors within the instructor's sphere of influence, while only 3% discussed the size of the course. This finding aligns with studies of York *et al.* and Walker *et al.*, suggesting that students perceive course-level inclusion as encompassing elements such as (i) providing resources, (ii) having a well-structured course that facilitates learning through various modalities, (iii) providing opportunities for interaction with the instructor, and (iv) setting up opportunities for engaging with peers who share similar questions. It is important to note that the study of Walker *et al.* ($N = 550$) was conducted in a general chemistry 1 (GC1) classroom at the same institution, as the current study, undergoing similar course changes as IP1 due to the Covid-19 pandemic [33]. However, the study of York *et al.* ($N = 1928$) was conducted in five large introductory STEM courses over one academic year (introductory biology I, general chemistry I and II, and introductory physics I and II) at a private institution prior to the Covid-19 pandemic [34]. Despite differences in institution, class size, class demographics, discipline, and classroom setting, most student responses were characterized by an academic factor (i.e., course structure, environment, student-student interaction, or student-instructor interaction) to describe their perception of course-level inclusion. More importantly, the same factors that emerged from student responses to construct the academic subcategories align with the findings of Singh, Li, and Cwick [25–27,60]. That is, inclusive learning environments, from a student's perspective, include factors such as interaction with peers (student-student interaction), how students interact with their instructor, the structure of the course, and what the classroom environment is like.

From student voices, it seems that the instructor's ability to make IP1 interactive and collaborative seems to override class size (see Table VII, for example, quotes from the academic subcategories). Our data suggest that instructors have influence over numerous factors that students identify as impacting their course inclusion in IP1. Examples of improving course inclusion include incorporating small-group work during class, including involving the use of peer leaders such as LAs or Peer-led Team Learning

TABLE VII. Example quotes from academic subcategories.

	Example quotes
Course structure—teaching practices	"This course was very flexible in the ways it offered help and resources to students and allowed many choices in approaching situations in the class. I thought that allowing some students to work alone in the discussion sessions was a good idea, since it allows a personal choice of how one can approach the coursework."
	"There is much encouragement to interact with other students and TAs whenever you want or need."
Environment	"The teaching team is very good about catering to everyone's needs, especially taking into account current events. I felt that I was cared about and catered to, which is really nice."
	"I felt heard and seen and loved the consistent encouragement from the learning team"
Student-instructor interaction	"The instructor is very encouraging and welcomes class participation."
	"It's a good course and I feel like my questions are welcomed by the teaching team"

(PLTL) leaders, which fosters student discussions and collaborative learning. Previous studies have shown that incorporating learning assistants (LAs) into the course structure improved student understanding of basic content knowledge in physics [61], bridged the communication gap between students and their instructor [62], and removed learning gaps [63]. Moreover, Clements *et al.* found that student responses indicated that LAs promoted a sense of belonging in STEM courses by decreasing feelings of isolation, serving as inspirational role models, and clarifying progression through the STEM educational system [64]. Additionally, incorporating PLTL pedagogy has been shown to increase performance and retention in STEM classrooms [65]. More so, PLTL methods have been shown to be a helpful pedagogy for underrepresented groups in STEM, including women and students from certain racial or ethnic marginalized groups [65–69]. Additional recommendations for instructors involve (i) offering various communication channels for students to engage with you, such as virtual and in-person office hours, and providing clear explanations of the instructor's purposes for office hours (or "student hours") [70]; (ii) engaging in informal one-on-one conversations with students before and after class; (iii) fostering a culture of growth mindset that emphasizes the importance of process and improvement [23]; (iv) monitoring student participation to ensure a diverse range of voices is encouraged [71]; and (v) establishing an atmosphere characterized by respect, openness, and community [72,73]. Instructors should also be mindful when interacting with students. For example, using words such as "simple," "easy," or "obvious" can discourage students and disparage students who already face stereotypes in physics [25]. Instead, instructors can improve their course learning environment by adopting culturally sensitive teaching methods and supporting underrepresented students through mentoring [74].

B. Students who did not feel IP1 was inclusive attributed their reasons to remote learning factors

We observed distinct factors influencing students' perceptions of course-level inclusion in different self-reported course-inclusion groups—high and low. The low inclusion group comprised students who gave a Likert-scale rating of (1) not at all, (2) slightly, or (3) somewhat inclusive ($n = 32$), while the high inclusion group included those who rated it as (5) highly inclusive ($n = 172$). In the group of students reporting low course inclusivity, their primary reasons related to remote learning. However, while there is a difference between the high or low inclusivity score groups within the primary categories, we could not connect a specific RL subcategory factor to those who had reported a low inclusion score due to a small sample size. While this limitation exists, our findings align with the work of Walker *et al.* on a remote learning general chemistry 1 course during the Covid-19 pandemic. In their study, Walker *et al.* found that there was a difference in distribution in how students in the remote-learning subcategory used different factors to describe their course-level inclusivity. However, after applying a Bonferroni correction, they could not attribute this difference in response distribution to any specific remote-learning subcategory [33].

In addition, for all students whose responses were coded in the primary remote learning category, from our data, we found that students who attributed their inclusivity to remote-learning factors tended to be nonspecific in their responses (RL-nonspecific subcategory, 39%, $n = 20$) or additionally mentioned course structure (RL-course structure, 33%, $n = 17$) in their responses. Responses in the RL-nonspecific subcategory predominantly reported that instructors made efforts to create an inclusive course. However, due to the course structure enforced by the pandemic, students felt that nothing more could be done to improve the inclusivity of the class, thus attributing their main reasons for their Likert-scale response to an overall remote-learning effect. The following are example responses: "Almost everything was done to make the class more inclusive, the teaching team did a good job, but the online format is really not preferred."; "Because it was an online course, I feel like this dampened the effects."; "I feel like it is hard to be included with the purely Zoom online classes." Example quotes of RL-course structure included, "This class has been very inclusive with breakout rooms and many opportunities to ask questions and work with others." and "It is hard to be fully inclusive in online learning and the job falls more on the student."

The online environment of IP1 courses during the Covid-19 pandemic stirred mixed feelings among students. Across the nation, the majority of students in introductory physics courses expressed concern about drastically reduced social interactions [75]. According to Moore, the communication aspects of an online course involve three types of interactions: student-student, student-content, and student-instructor [37]. Implications for instructors in online environments involve emphasizing "active student engagement" even more so than in courses that are face-to-face [76,77]. For instructors new to fostering active student engagement in online settings, it is advisable to start by introducing a limited number of constructive and interactive activities that align closely with the instructor's learning objectives. These can be gradually expanded upon in subsequent course offerings [78]. Petersen's recommendations for establishing an inclusive learning environment include designing course activities that are consistent with the course's credit hours, providing an informative syllabus, and fostering an environment where students understand the course expectations [41]. A comprehensive syllabus should outline major topics and allocate course time, learning outcomes, and objectives [78]. Each topic should be clearly defined, with corresponding content and learning objectives, while ensuring that extraneous material is trimmed down. Additionally, it is essential to plan engaging activities that allow students to practice and receive feedback [78].

C. Identity's potential impact on perceptions of inclusivity

In the current study, the identity category emerged from student quotes that highlighted the significance of a person's identity in shaping their level of engagement in the learning process. This influence was conveyed through language, course content, or referred to students' previous knowledge. While the term "identity" carries nuanced interpretations and implications in literature, personal identity can be described by elements such as one's name, age, gender, race, ethnicity, and individual experiences. Historical context reveals that a predominance of men are enrolled in calculus-based college-level introductory physics courses [6]. Studies focusing on women have found that women's underrepresentation is attributed to persistent gender-science stereotypes [13,79]. Moreover, women face challenges in identifying with STEM fields due to signals from their social environments suggesting that women do not belong in the sciences [80]. Concurrently, past research demonstrates that Black women and women of color encounter barriers rooted in their racial identities, resulting in limited recognition and exclusion from physics [81–83]. However, in this study, only 3% ($n = 10$) of student responses involved factors discussing identity factors (four women and six men). Instead, many students, including women, used academic factors to describe their perception of course-level inclusion rather than identity. When looking closely at the women who discussed identity as a factor of their course-level inclusion, they referenced their instructors' awareness of pronouns in their responses:

> The professor seems like she tries to cater to the various learning styles that students may have. She is also attentive to the proper use of pronouns, which is a good thing. (Woman, White, first-generation student).

> I think the course has been very inclusive especially with such a care taken to call people the right name and pronouns. (Woman, White, non-first generation).

> Everyone got along really well and Prof [P2] was also very good with the use of pronouns and such to ensure that every student felt safe and comfortable (Woman, RM, first generation).

This finding stood out from previous studies done by Walker *et al.* and York *et al.* In the study of York *et al.*, the women who discussed identity as a factor of their course-level inclusion talked about the preknowledge requirement. In their study, they attributed this focus as a result of more students taking advanced placement (AP) courses in high school and therefore, students believed that AP was a part of preknowledge. In the work of Walker *et al.*, the students who used identity factors to discuss their perception of course-level inclusivity ($N = 11$) described a variety of identity-specific factors such as previous knowledge ($n = 6$), gender ($n = 2$), sexual orientation ($n = 1$), differences in ability ($n = 1$), and being a nontraditional student ($n = 1$). As with York *et al.*, in the study of Walker *et al.*, students who used identity factors to describe their course-level inclusivity were more likely to discuss "previous knowledge" as contributing to course-level inclusion. One possible reason for this difference in focus on previous knowledge could be that the GC1 classrooms studied by Walker *et al.* and all the classes in York *et al.* contained both health and physical science majors. Whereas in our study, IP1 was tailored for physical science majors who typically take calculus as part of their curriculum, while students majoring in the health sciences take an algebra-based introductory physics course (less quantitative).

Our findings on the percentage of students who discussed identity are consistent with the work of York *et al.* [34] in a nonremote learning environment before the Covid-19 pandemic and the study of Walker *et al.* [33] on a remote learning GC1 course during the Covid-19 pandemic, in that Identity was a small percentage relative to academic-factors majority, which comprised at least 40% of responses in all three studies. In the research of Walker *et al.*, 2% of students ($n = 11$) in GC1 ($N = 550$) used identity factors to describe their course-level inclusion, with the majority of those responses focusing on entering GC1 with or without prior knowledge. York *et al.* ($N = 1928$) saw that approximately 14% of students, which included IP1, introductory biology, and general chemistry, discussed identity factors concerning their perception of course-level inclusion. A possible contributing factor to the lower percentage of identity responses in the current study and the study of Walker *et al.*, compared to York *et al.*, could be the remote nature of the IP1 and GC1 courses. Unlike in-person learning environments, where students can see the composition of the classroom, in the remote-learning environment, the social identity of their peers is typically unknown because many students may have opted to turn off their cameras or remain muted during Zoom lectures. Additionally, many online platforms typically can only show a small percentage of students in the class.

In addition, although students could write multiple reasons for their level of inclusivity, all students wrote one main reason, and hence, identity could have been more important to students, as other literature has shown, than is seen in our results. It could also be that at this level, students might be less aware of the effect their identity has on their feelings of inclusivity, course performance, and retention. Instructors may support these students by implementing evidence-based interventions focused on increasing diverse identities in STEM, such as (i) using references and examples in class that include different identities (e.g., gender, race, ethnicity, generational status, etc.) during the course [71,84]; (ii) include scientist spotlights throughout the semester and have students write reflections [85,86]; and (iii) conduct social-psychological interventions such as value-affirmation tasks and social-belonging reflections [60,87].

VI. LIMITATIONS

In this study, we investigated students' perspectives on inclusion in an introductory physics 1 (IP1) calculus-based course at an R1 institution in the United States. While this exploratory study focused on listening to student voices and what they perceive affected their inclusion in this course, there are several limitations that must be discussed. First, this study took place during the Covid-19 pandemic, which was a period of great uncertainty, change, and stress for students. Due to the nature of the pandemic, many courses, such as IP1, were forced to become remote learning, leaving students with no in-person interaction with their peers or instructors. As a result, it may have impacted what factors students used to describe their course inclusion and may be different from the ones they might have used for an in-person IP1 course. Despite this unusual time and remote-learning format of IP1, the findings of this study are congruent with those of a similar study ($N = 1928$) that examined large, in-person introductory STEM courses prior to the pandemic. A second limitation was the use of highly simplified demographic variables; therefore, the nuances and complexities of gender identity, racial or ethnicity identity, race-gender intersections, and other identities and identity intersections were not captured in the results. Furthermore, while we examined the variation in the number of responses for each theme across groups, we did not qualitatively explore potential variations in the content of students' statements within each theme based on the group; both are important areas for future study. Third, a significant percentage of students (33%) gave nonspecific reasons for the level of inclusivity they perceived in IP1. Although these nonspecific reasons do not elucidate factors students use to describe inclusivity, our future work involves developing a quantitative course-level inclusion survey based on this qualitative work. We anticipate that this survey will yield more specific responses, contributing to a deeper understanding of the factors influencing students' perceptions of course inclusivity. York *et al.*'s study discusses the value of developing a preliminary Likert-scale survey; a pilot study of such a survey resulted in students giving at least one reason from a given list of factors. A fourth limitation of this study was the sample size. While different results could have been seen in an identical classroom of a larger size, similar trends found in the current study were also observed in the work of Walker *et al.* ($N = 550$) (general chemistry 1 course at the same public R1 institution) and the work of York *et al.* ($N = 1928$) (several introductory STEM courses in a select private institution). Finally, students predominantly gave a single reason for their perceived inclusivity level. It is reasonable to assume that students consider multiple factors, but the open-ended option may have encouraged them to provide only one. Once again, a quantitative course-inclusion survey is expected to reveal the nuanced reasons students employ when evaluating their course-level inclusion.

VII. CONCLUSION

In this exploratory study of a calculus-based introductory physics 1 (IP1) ($N = 289$), which was a remote course as a result of the Covid-19 pandemic, we examined students' perceptions of course inclusion using a Likert scale. The mean rating for the inclusivity of IP1 was "moderately inclusive" ($M = 4.41$, $\mathrm{SD} = 0.79$). Qualitative analysis of open-ended responses revealed six primary categories, with academic and nonspecific comprising 41% and 33% of all responses, respectively, and remote learning being 18%. Subcategories were identified within academic and remote learning. Our results showed that the majority of the students described academic factors (i.e., classroom practices and structures) that were within the instructor's influence. Chi-square tests indicated that students perceiving high inclusion emphasized academic factors, while those feeling low inclusion focused on remote learning. These findings emphasize the need for inclusive practices, reflecting student preferences for interaction with instructors and peers and a welcoming atmosphere. Additional studies with larger sample sizes for in-person and online approaches are needed to further understand what factors students use in determining their inclusion in these introductory STEM courses and the variations among different subgroups.

ACKNOWLEDGMENTS

We would like to thank the instructors (Ramón S. Barthelemy and Claudia De Grandi) and the students of IP1 included in this study. Ashton M. Barber assisted in the initial coding of the remote learning quotes. H. E. B. was supported by Undergraduate Research Opportunity Program (UROP) funds from the University of Utah.

[1] S. Cwik and C. Singh, Students' sense of belonging in introductory physics course for bioscience majors predicts their grade, Phys. Rev. Phys. Educ. Res. **18**, 010139 (2022).

[2] J. D. Edwards, L. Laguerre, R. S. Barthelemy, C. De Grandi, and R. F. Frey, Relating students' social belonging and course performance across multiple assessment types in a calculus-based introductory physics 1 course, Phys. Rev. Phys. Educ. Res. **18**, 020150 (2022).

[3] B. Ferrell, M. M. Phillips, and J. Barbera, Connecting achievement motivation to performance in general chemistry, Chem. Educ. Res. Pract. **17**, 1054 (2016).

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

[5] Y. Wang, G. A. Rocabado, J. E. Lewis, and S. E. Lewis, Prompts to promote success: Evaluating utility value and growth mindset interventions on general chemistry students' attitude and academic performance, J. Chem. Educ. **98**, 1476 (2021).

[6] National Center for Science and Engineering Statistics (NCSES), Diversity and STEM: Women, Minorities, and Persons with Disabilities 2023, Special Report NSF 23-315 (National Science Foundation, 2023), https://ncses.nsf.gov/wmpd.

[7] S. Cwik and C. Singh, How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses, Phys. Rev. Phys. Educ. Res. **17**, 020143 (2021).

[8] E. Marshman, Z. Y. Kalender, C. Schunn, T. Nokes-Malach, and C. Singh, A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences, Canadian J. Phys. **96**, 391 (2017).

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

[10] D. Doucette and C. Singh, Why are there so few women in physics? Reflections on the experiences of two women, Phys. Teach. **58**, 297 (2020).

[11] Z. Hazari, E. Brewe, R. M. Goertzen, and T. Hodapp, The importance of high school physics teachers for female students' physics identity and persistence, Phys. Teach. **55**, 96 (2017).

[12] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way, Phys. Rev. Phys. Educ. Res. **15**, 020148 (2019).

[13] A. Maries, N. I. Karim, and C. Singh, Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics? Phys. Rev. Phys. Educ. Res. **14**, 020119 (2018).

[14] K. Rosa and F. M. Mensah, Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life, Phys. Rev. Phys. Educ. Res. **12**, 020113 (2016).

[15] Z. Y. Kalender, E. Marshman, T. J. Nokes-Malach, C. Schunn, and C. Singh, Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses, presented at PER Conf. 2017, 10.1119/perc.2017.pr.046.

[16] C. Moreno, D. Pham, and L. Ye, Chemistry self-efficacy in lower-division chemistry courses: Changes after a semester of instruction and gaps still remain between student groups, Chem. Educ. Res. Pract. **22**, 772 (2021).

[17] J. D. Edwards, R. S. Barthelemy, and R. F. Frey, Relationship between course-level social belonging (sense of belonging and belonging uncertainty) and academic performance in General Chemistry 1, J. Chem. Educ. **99**, 71 (2021).

[18] A. Fink, R. F. Frey, and E. D. Solomon, Belonging in general chemistry predicts first-year undergraduates' performance and attrition, Chem. Educ. Res. Pract. **21**, 1042 (2020).

[19] K. L. Lewis, J. G. Stout, S. J. Pollock, N. D. Finkelstein, and T. A. Ito, Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics, Phys. Rev. Phys. Educ. Res. **12**, 020110 (2016).

[20] J. R. Hoehn and N. D. Finkelstein, Fostering inclusion and teaching equity in a Modern Physics for Engineers course, in *Proceedings of the 2020 ASEE Virtual Annual Conference Content Access* (2020), 10.18260/1-2–34684.

[21] B. Dewsbury and C. J. Brame, Inclusive teaching, CBE Life Sci. Educ. **18**, fe2 (2019).

[22] M. d. C. Salazar, A. S. Norton, and F. A. Tuitt, 12: Weaving promising practices for inclusive excellence into the higher education classroom, To improve the academy **28**, 208 (2010).

[23] K. N. White, K. Vincent-Layton, and B. Villarreal, Equitable and inclusive practices designed to reduce equity gaps in undergraduate chemistry courses, J. Chem. Educ. **98**, 330 (2021).

[24] J. Handelsman, S. Elgin, M. Estrada, S. Hays, T. Johnson, S. Miller, V. Mingo, C. Shaffer, and J. Williams, Achieving STEM diversity: Fix the classrooms, Science **376**, 1057 (2022).

[25] S. Cwik and C. Singh, Role of inclusiveness of learning environment in predicting students' outcomes in courses in which women are not underrepresented, J. Higher Educ. Theory Pract. **22**, 176 (2022).

[26] Y. Li and C. Singh, Inclusive learning environments can improve student learning and motivational beliefs, Phys. Rev. Phys. Educ. Res. **18**, 020147 (2022).

[27] Y. Li and C. Singh, How inclusiveness of learning environment predicts female and male students' physics grades and motivational beliefs in introductory physics courses, presented at PER Conf. 2022, 10.1119/perc.2022.pr.Li.

[28] G. M. Walton and G. L. Cohen, A question of belonging: Race, social fit, and achievement, J. Pers. Soc. Psychol. **92**, 82 (2007).

[29] Y. Li and C. Singh, Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses, Phys. Rev. Phys. Educ. Res. **17**, 010143 (2021).

[30] R. Masika and J. Jones, Building student belonging and engagement: Insights into higher education students' experiences of participating and learning together, Teach. Higher Educ. **21**, 138 (2016).

[31] M. Meeuwisse, S. E. Severiens, and M. P. Born, Learning environment, interaction, sense of belonging and study success in ethnically diverse student groups, Res. High. Educ. **51**, 528 (2010).

[32] B. M. Dewsbury, H. J. Swanson, S. Moseman-Valtierra, and J. Caulkins, Inclusive and active pedagogies reduce academic outcome gaps and improve long-term performance, PLoS One **17**, e0268620 (2022).

[33] D. K. Walker, V. P. Bustamante, H. E. Blomgren, J. D. Edwards, and R. F. Frey, A mixed methods study exploring students' open-ended responses about course inclusivity in General Chemistry 1, J. Chem. Educ. **101**, 1403 (2024).

[34] A. M. York, K. G. Miller, M. J. Cahilll, M. A. Bernstein, A. M. Barber, H. E. Blomgren, and R. F. Frey, What factors contribute to students' perceptions of inclusion in introductory STEM courses: A mixed-methods exploratory study, CBE Life Sci. Educ. (to be published).

[35] L. A. Arasaratnam-Smith and M. T. Northcote, Community in online higher education: Challenges and opportunities, Electron. J. e-Learning **15**, 188 (2017), retrieved from www.ejel. org, https://files.eric.ed.gov/fulltext/EJ1141773.pdf.

[36] N. P. Wingo, N. V. Ivankova, and J. A. Moss, Faculty perceptions about teaching online: Exploring the literature using the technology acceptance model as an organizing framework, Online Learn. **21**, 15 (2017).

[37] M. G. Moore, Editorial: Three types of interaction, Am. J. Distance Educ. **3**, 1 (1989).

[38] L. M. Archambault and J. H. Barnett, Revisiting technological pedagogical content knowledge: Exploring the TPACK framework, Comput. Educ. **55**, 1656 (2010).

[39] S. Cox and C. R. Graham, Diagramming TPACK in practice: Using an elaborated model of the TPACK framework to analyze and depict teacher knowledge, TechTrends : for leaders in education and training **53**, 60 (2009).

[40] M. Sailer, F. Schultz-Pernice, and F. Fischer, Contextual facilitators for learning activities involving technology in higher education: The C♭-model, Comput. Hum. Behav. **121**, 106794 (2021).

[41] N. J. Petersen, Strategies for efficient, meaningful, and inclusive online learning environments: It's about time, in *Handbook of Research on Creating Meaningful Experiences in Online Courses* (IGI Global, Hershey, PA, 2020), p. 187.

[42] A. P. Barrasso and K. E. Spilios, A scoping review of literature assessing the impact of the learning assistant model, Int. J. STEM Educ. **8**, 1 (2021).

[43] Qualtrics, Provo, UT, 2020, https://www.qualtrics.com.

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.20.020112 for descriptive statistics for course-inclusivity rating include a histogram of student ratings and the percentage and number of students in each rating score; detailed codebook; interrater reliability (IRR) values; statistical results for primary categories, academic subcategories, and remote-learning subcategories.

[45] G. A. Chen and J. Y. Buell, Of models and myths: Asian (Americans) in STEM and the neoliberal racial project, Race Ethn. Educ. **21**, 607 (2018).

[46] E. O. McGee, B. K. Thakore, and S. S. LaBlance, The burden of being "model": Racialized experiences of Asian STEM college students, J. Diversity Higher Educ. **10**, 253 (2017).

[47] D. J. Asai, Race matters, Cell **181**, 754 (2020).

[48] R. Krakehl and A. M. Kelly, Intersectional analysis of Advanced Placement Physics participation and performance by gender and ethnicity, Phys. Rev. Phys. Educ. Res. **17**, 020105 (2021).

[49] J. D. Edwards, H. L. Torres, and R. F. Frey, The effect of social belonging on persistence to general chemistry 2., J. Chem. Educ. **100**, 4190 (2023).

[50] J. Y. Cho and E.-H. Lee, Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences, Qual. Rep. **19**, 1 (2014).

[51] H.-F. Hsieh and S. E. Shannon, Three approaches to qualitative content analysis, Qual. Health Res. **15**, 1277 (2005).

[52] S. B. Merriam, *Qualitative Research: A Guide to Design and Implementation* (Jossey-Bass Inc. Pub, San Francisco, CA, 2009).

[53] D. Freelon, ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service, Int. J. Internet Sci. **8**, 10 (2013), https://dfreelon.org/publications/2013_ReCal_OIR_Ordinal_Interval_and_Ratio_Intercoder_Reliability_as_a_Web_Service.pdf.

[54] ReCal OIR, https://dfreelon.org/utils/recalfront/recal-oir/ [accessed April 24, 2023].

[55] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology* (Sage Publications, Thousand Oaks, CA, 2018).

[56] H.-Y. Kim, Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test, Restor. Dent. Endod. **42**, 152 (2017).

[57] R. J. Feise, Do multiple outcome measures require p-value adjustment? BMC Med. Res. Methodol. **2**, 1 (2002).

[58] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge, New York, 1988).

[59] STATA Statistical Software: Release 17, College Station, TX, 2019.

[60] C. Singh, Using social psychological interventions to make physics classes equitable and inclusive, arXiv:2112.10732.

[61] V. Otero, S. Pollock, and N. Finkelstein, A physics department's role in preparing physics teachers: The Colorado learning assistant model, Am. J. Phys. **78**, 1218 (2010).

[62] P. C. Hamerski, P. W. Irving, and D. McPadden, Learning assistants as student partners in introductory physics, Phys. Rev. Phys. Educ. Res. **17**, 020107 (2021).

[63] B. Van Dusen, J.-S. S. White, and E. Roualdes, The impact of learning assistants on inequities in physics student outcomes, arXiv:1607.07121.

[64] T. P. Clements, K. L. Friedman, H. J. Johnson, C. J. Meier, J. Watkins, A. J. Brockman, and C. J. Brame, "It made me feel like a bigger part of the STEM community": Incorporation of learning assistants enhances students' sense of

belonging in a large introductory biology course, CBE Life Sci. Educ. **21**, ar26 (2022).

[65] D. Drane, H. D. Smith, G. Light, L. Pinto, and S. Swarat, The gateway science workshop program: Enhancing student performance and retention in the sciences through peer-facilitated discussion, J. Sci. Educ. Technol. **14**, 337 (2005).

[66] R. F. Frey, A. Fink, M. J. Cahill, M. A. McDaniel, and E. D. Solomon, Peer-led team learning in general chemistry I: Interactions with identity, academic preparation, and a course-based intervention, J. Chem. Educ. **95**, 2103 (2018).

[67] S. Horwitz *et al.*, Using peer-led team learning to increase participation and success of under-represented groups in introductory computer science, ACM SIGCSE Bull. **41**, 163 (2009).

[68] S. E. Lewis, Retention and reform: An evaluation of peer-led team learning, J. Chem. Educ. **88**, 703 (2011).

[69] S. B. Wilson and P. Varma-Nelson, Small groups, significant impact: A review of peer-led team learning research with implications for STEM education researchers and faculty, J. Chem. Educ. **93**, 1686 (2016).

[70] M. Smith, Y. Chen, R. Berndtson, K. M. Burson, and W. Griffin, "Office hours are kind of weird": Reclaiming a resource to foster student-faculty interaction, InSight **12**, 14 (2017), https://files.eric.ed.gov/fulltext/EJ1152098.pdf.

[71] K. D. Tanner, Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity, CBE Life Sci. Educ. **12**, 322 (2013).

[72] M. Estrada, P. R. Hernandez, and P. W. Schultz, A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers, CBE Life Sci. Educ. **17**, ar9 (2018).

[73] K. Rainey, M. Dancy, R. Mickelson, E. Stearns, and S. Moller, Race and gender differences in how sense of belonging influences decisions to major in STEM, Int. J. STEM Educ. **5**, 1 (2018).

[74] T. Dobbins, The time is now: Systemic changes to increase African Americans with bachelor's degrees in physics and astronomy, American Institute of Physics: TEAM-UP Report, 2020, https://www.aip.org/sites/default/files/aipcorp/files/teamup-full-report.pdf.

[75] M. Dew, L. Ford, D. T. Nodurft, T. Erukhimova, and J. Perry, Student responses to changes in introductory physics learning due to the COVID-19 pandemic, Phys. Teach. **59**, 162 (2021).

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

[77] E. J. Theobald *et al.*, Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math, Proc. Natl. Acad. Sci. U.S.A. **117**, 6476 (2020).

[78] M. Prince, R. Felder, and R. Brent, Active student engagement in online STEM classes: Approaches and recommendations, Adv. Eng. Educ. **8**, 1 (2020), https://advances.asee.org/wp-content/uploads/Covid%2019%20Issue/Text/2%20AEE-COVID-19-Felder.pdf.

[79] M. Dew, J. Perry, L. Ford, W. Bassichis, and T. Erukhimova, Gendered performance differences in introductory physics: A study from a large land-grant university, Phys. Rev. Phys. Educ. Res. **17**, 010106 (2021).

[80] A. Y. Kim, G. M. Sinatra, and V. Seyranian, Developing a STEM identity among young women: A social identity perspective, Rev. Educ. Res. **88**, 589 (2018).

[81] S. Hyater-Adams, C. Fracchiolla, T. Williams, N. Finkelstein, and K. Hinko, Deconstructing Black physics identity: Linking individual and social constructs using the critical physics identity framework, Phys. Rev. Phys. Educ. Res. **15**, 020115 (2019).

[82] M. Ong, Body projects of young women of color in physics: Intersections of gender, race, and science, Soc. Probl. **52**, 593 (2005).

[83] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).

[84] S. Cheryan, A. N. Meltzoff, and S. Kim, Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes, Comput. Educ. **57**, 1825 (2011).

[85] J. N. Schinske, H. Perkins, A. Snyder, and M. Wyer, Scientist spotlight homework assignments shift students' stereotypes of scientists and enhance science identity in a diverse introductory science class, CBE Life Sci. Educ. **15**, ar47 (2016).

[86] K. J. Metzger, M. Dingel, and E. Brown, "No matter what your story is, there is a place for you in science": Students' ability to relate to scientists positively shifts after scientist spotlight assignments, especially for first-generation students and women, CBE Life Sci. Educ. **22**, ar12 (2023).

[87] G. M. Walton *et al.*, Where and with whom does a brief social-belonging intervention promote progress in college? Science **380**, 499 (2023).