Supplemental Material

CBE—Life Sciences Education

York *et al*.

Supplemental Materials

An exploratory mixed-methods analysis of factors contributing to students' perceptions of inclusion in introductory STEM courses

Alessandra M. York[a,b], Kathryn G. Miller[b], Michael J. Cahill[a], Mindy A. Bernstein[c], Ashton M. Barber[d], Hannah E. Blomgren[e] and Regina F. Frey[e]*

Supplemental Materials

A. Additional Course Information: Class Sizes

Class Sizes

- Introductory Biology 1-Spring Semester- total N=623; Section 1= 357, Section 2=266
- General Chemistry
    - GC1-Fall Semester- total N=705; Section 1=132, Section 2=347, Section 3=226
    - GC2-Spring Semester- total N=552; Section 1= 211, Section 2= 341
- Introductory Physics
    - IP1-Fall Semester- total N=629; Section 1=102, Section 2=107, Section 3=108, Section 4=106, Section 5=114, Section 6=92
    - IP2-Spring Semester- total N=542; Section 1=52, Section 2=110, Section 3=114, Section 4=73, Section 5=112, Section 6=81

B. Additional Course Information: Course Descriptions

Most of the courses had three one-hour in-person lectures (physics had some sections that were two 80-min in-person lectures); all the courses had some active-learning components during their lectures with physics having the greatest amount of active learning (York et al., 2021), and all had some type of supplemental collaborative-learning problem-solving program such as Peer-Led Team Learning (PLTL) (Frey & Lewis, 2023). While biology and physics had associated laboratory components with their courses (chemistry had a separate laboratory course), the survey asked the students to only consider the non-laboratory components for their responses.

At this institution, a typical life sciences/pre-health major takes General Chemistry 1 in their first semester in the fall, and concurrently enrolls in General Chemistry 2 and Introductory Biology 1 in the following spring. These students then take Introductory Physics 1 and 2 in their third year of school. See Figure 1 in the main text to see the flow of students between courses.

Introductory Biology Course

During this time period, Introductory Biology 1 was the first course in the introductory biology sequence and was offered in the spring semester. Approximately 700 mostly first-year students enrolled in the course. Many were concurrently enrolled in the second semester of general chemistry. Introductory Biology 1 had 2 sections and two instructors. Each instructor taught approximately half of the lectures in both sections. There were three one-hour in-person lectures and an associated lab component (2.5 hours each week). Assessments included weekly quizzes, iClicker participation, three exams and a cumulative final. Supplemental learning opportunities included instructor and teaching-

assistant office hours as well as an optional Biology Team Learning (BTL) program, in which older undergraduate peers facilitated small groups of students in solving/discussing the weekly homework sets.

General Chemistry Courses

During this study, approximately 800 students enrolled in General Chemistry 1 in the fall semester, with approximately 84% first year students. All pre-STEM majors including biology, chemistry, physics, pre-health and engineering majors took this course. With 3 sections (200-350 students/section), and a team of 5 instructors (3 gave lectures), General Chemistry 1 comprised 3 one-hour in-person lectures with weekly recitation classes consisting of students from multiple sections. Although General Chemistry 1 had multiple sections with different instructors, the course was treated as one course with the same syllabus and policies, clicker questions, graded and ungraded homework, weekly quizzes, three exams, along with a cumulative final exam, access to PLTL packets, and combined absolute grading. Students were offered a variety of supplemental programs that included daily optional help sessions held by faculty members in the instructor team, optional enrollment in the PLTL (Peer-Led Team Learning) program (Frey, Fink, Cahill, McDaniel, & Solomon, 2018; Hockings, DeAngelis, & Frey, 2008), which provided facilitated problem-solving practice in a group-study environment following the PLTL approach with trained peer leaders, and the General Chemistry Transition Program (Shields et al., 2012) for students who scored in the bottom 25% on the department's chemistry online diagnostic exam.

General Chemistry 2 (GC2) was the second course in the general chemistry series offered in the spring semester with approximately 500-600 students enrolled; General Chemistry 2 was smaller because some engineering majors only take the first semester of General Chemistry. There were 4 instructors in the course (2 gave lectures). General Chemistry 2 was structured similar to the first semester (i.e., treated as one large course with the same assignments, structure, policies, and supplemental programs). For both, the laboratory component was a separate course, in which students concurrently enrolled.

Introductory Physics Courses

Approximately 600-650 students enrolled in the first course of the introductory calculus-based physics sequence during this study. Similar to the general chemistry sequence above, the students were composed of the same STEM majors or tracks with the addition of architecture majors. Unlike the other introductory STEM courses, which comprised mostly first-year students, introductory physics had similar

percentages of first-year (43%) and continuing students (57%). Introductory Physics 1 used a flipped-classroom approach in which students watched videos prior to attending lecture; therefore, there was a greater focus on working through examples and problem-solving in small groups during the lecture relative to other introductory STEM courses (York et al., 2021). There were five instructors offering six course sections of three 50-minute in-person lectures or two 80-minute in-person lectures with a weekly lab. Students from all six lecture sections (120 or less students/section) took the same three midterm exams, graded and ungraded assignments, and labs, with PLTL offered as an optional supplemental resource. Instructors also give demonstrations on a regular basis. Students also attend one three-hour lab each week, which contributes 20% of the final grade. Three exams are administered in Introductory Physics 1, each of which is worth 19% of the grade (57% total from exams). The rest of the final grade is determined by iClicker activities (3%), completion of the FlipItPhysics lectures (5%), and weekly homework assignments through Mastering Physics (15%). Course instructors held office hours so that students could ask questions. Academic mentoring through the university's academic help center (Cornerstone) was also offered. If students wanted more help, they were permitted to pay for private tutoring through the physics department. Introductory Physics 1 covered topics in mechanics including the laws governing motion and the theory of relativity.

Introductory Physics 2 was the second course offered in this introductory physics sequence and enrolled around 500-550 students. Most students from the first semester continued to take the second semester in the spring semester, except for architecture students. Four of the five instructors from Introductory Physics 1 continued to teach the six lecture sections. Introductory Physics 2 had the same course structure as the first semester (see above). It also has a co-requisite of Calculus II course. The content focuses on electricity and magnetism with a brief introduction into nuclear physics.

C. Interrater Reliability Chart

	Categories/Subcategories	**Interrater Reliability (α)**
Primary Categories	**Academic**	**0.872**
	Identity	**0.906**
	Nonspecific	**0.852**
	Do not value inclusion	**---**
	STEM is objective	**---**
	Other	**---**
Academic Subcategories	**Environment**	**0.476**
	Student-Instructor Interaction	**0.729**
	Course Structure	**0.672**

D. Descriptive Statistics for the Overall Sample Population

Supplemental Table 1

Class	Average Inclusion Score (N)	Number of Students Scoring "Not at all" (1)	Number of Students Scoring "Slightly" (2)	Number of Students Scoring "Somewhat" (3)	Number of Students Scoring "Moderately" (4)	Number of Students Scoring "Highly" (5)	SD
Class 1 (Spring 2019)	**3.69, N=412**	**15**	**45**	**98**	**146**	**108**	**1.08**
Class 2 (Fall 2018)	**4.12, N=370**	**3**	**20**	**58**	**138**	**151**	**0.92**
Class 3 (Spring 2019)	**4.09, N=289**	**4**	**13**	**48**	**111**	**113**	**0.93**
Class 4 (Fall 2018)	**4.02, N=548**	**11**	**34**	**90**	**210**	**203**	**0.98**
Class 5 (Spring 2019)	**4.28, N=451**	**4**	**11**	**53**	**170**	**213**	**0.83**

E. Overview of Codebook

Example Quotes from Each Category and Subcategory[a].

Primary Category	Subcategories	Example Quotes
Academic	*Environment	"The professors made an effort to ensure that both people from the front, sides, and back of the classroom could hear and engage with the material."
	*Student-Instructor Interaction	"professor encouraged lots of interacting and made sure to mention that anything asked or thought is fine. I felt even if I said the wrong answer no one would judge"

	*Course Structure	"I really enjoyed the clicker questions. It gave the students a voice and a way to interact during class."
Identity	*Professor Language	"I was impressed at how Dr [] includes their preferred pronouns in their email signature. It's a small gesture, but can go a long way for some students"
	*Content	"The professors made sure to point out contributions made by women when it pertained to the class."
	*Demographics /Background	"As a female I felt like the minority a lot of the time which feeds into my feelings of incompetence in the course."
	*Knowledge Background	"Some material is expected of students to know already without it being taught. This leaves out students who may have not come from strong high school backgrounds that did not efficiently teach these expected [subject] concepts."
STEM is objective	N/A	"There is honestly nothing that is not inclusive about the biology lecture. After all, it is a biology class where most information is factual and objective."
Do not value inclusion	N/A	"Too forced, it felt like they were going out of their way to make sure all the white men knew that are overrated, at one point she corrected a clipart silhouette of a white male doctor with a black woman. It is ridiculous, focus less on diversity and inclusion and more on actually teaching."
Other	N/A	"Since we only had lab lecture once a week for one hour, I felt that it was not enough time to fully understand what we needed to in class…"

[a]These are quotes in their entirety except for the quote in 'Other'. On average, the quotes typically were the length of a phrase or a sentence.

F. Coding Rules and Detailed Primary Category Definitions

Rules:

1) Responses should be coded with one category.

2) If the response could be coded with two, please choose the more specific response within the quote.

3) If the response is completely split between two categories, two can be coded but this should only happen if the topics are independent and/or unrelated.

4) Hierarchy is academic or identity, then STEM is objective/do not value inclusion, and last nonspecific. Always try to make it more specific if you can.

5) If they say one thing correct about inclusion that could be coded into academic or identity, then do NOT code STEM is objective or do not value inclusion, use either the academic or identity inclusion.

6) Do not code suggestions, unless the example has provided context to something that is occurring in the classroom related to either academic or identity inclusion.

7) Within identity inclusion, you may code both self-reported identity and another demographic identity subcode.

8) Within academic inclusion, attempt to subcode either participation or course structure before putting the quote in the Environment subcategory.

Nonspecific- Student described their experience of inclusion in the course, but gave no specific details. Ex. “It was inclusive. Nothing stood out to be wrong”

Academic- Student referenced teaching practices and course design elements that encourage all students to engage and feel a part of the learning process. This is done through creating an inclusive environment, creating opportunities for everyone to participate, and using evidence-based teaching practices. Ex. “My professor is a great at keeping the class engaged. He is able to get a lot of participation.” More quotes can be seen below in the Subcategory and theme definitions.

Identity- Student refer to a person’s identity and how it plays a role in encouraging all students to feel involved in the learning process. This is done through language, course content, or in a capacity related to the individuals involved in the course. Ex. “My professor made a few sexist comments and generally didn't make an inclusive environment”. More quotes can be seen below in the Subcategory and theme definitions.

STEM is Objective- Students expressed ideas that show they do not understand how STEM can incorporate inclusion and equity within the course. Ex. "There is honestly nothing that is not inclusive about the biology lecture. After all, it is a biology class where most information is factual and objective."

Do Not Value Inclusion- Student conveys that diversity, equity, and inclusion are not relevant or were overemphasized in the course. Ex. “There was a clear effort made to be "inclusive" but it was way too off-putting. I just want to learn about biology, I just want to learn about the information. Inclusivity is not important to me in the context of my bio lecture. I even felt that some of my learning was sacrificed in an effort to express diversity”

Other- The student did not discuss their classroom experience or they discussed a different class. Ex. “Discussing cellular biology is truly interesting, but I want to know more about how other kinds of life function [...]”

G. Detailed Subcategory and Theme Definitions

Subcategory Definitions

Academic

Environment- Student 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. Ex. “I find that I've made many friends through the Gen Chem classes, and we are all interested in the material, making for a more enjoyable learning experience.” “When we would have clicker questions, it would be very inclusive and facilitated a lot of discussion amongst my peers. It would feel like 1 homogenous class, not like we were all separate people just there and listening to one person talk.”

Student-Instructor Interaction- Student 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. Ex. “Dr. [] asked us questions in class to make sure we understood the topics and during the clicker, she walked around to help us which was very inclusive in helping us learn.”

Course Structure- Student referenced the aspects of the course design or infrastructure that affected the inclusivity. Ex. “I really appreciated my professor's understanding of my needs during my injury.”

Identity

Professor Language- Student addressed how inclusive the instructor’s language was when referencing a group with a certain identity. Ex. “I was impressed at how Dr [] includes her preferred pronouns in her email signature. It's a small gesture, but can go a long way for some students”

Culturally-relevant material- Student addressed the inclusivity of the course material, such as problems/questions and content on slides or in examples. Ex. “They made sure that they showed that a diverse community can be involved in the field of biology”

Demographics/Background- Student addressed aspects of demographics/background of individuals within the course or among the instructor team. Ex. “There is a large variety of students in my class. Many are like me and many are very different from me, but everyone participates equally”

Knowledge Background- Student addressed the level of previous knowledge they believed was necessary to be successful in the course, or any assumptions made by an instructor about their prior knowledge pertaining to the course (e.g., does or does not assume students have taken AP/IB/general course). Ex. “It was definitely geared towards people who had taken upper level high school physics courses such as AP Physics”

Subcategory Thread Definitions

Academic

Course Structure [Instructor-Controlled]- Student referenced the specific teaching practices used in class, the course policies, the prerequisites, the resources offered, the range/flow of the material covered by the course, or how well the material is explained by the instructor. Ex. “I really enjoyed the clicker questions. It gave the students a voice and a way to interact during class.”

*Course Structure [University-Controlled]-*Student referenced the size of the lecture or the physical accessibility of the lecture space. Ex. “Given the sizes of the classes, I feel that the professors always try their best to be as inclusive as possible, but it definitely is hard to feel like you're being "focused" on in large lectures[…]”

Identity

Demographics/Background [General]- Student addressed diversity of individuals within the course or of the instructor team. Ex. “The instructor did a good job of making the environment of physics lecture inclusive to all students. He did not judge any students based on their background in physics or personal characteristics.”

Demographics/Background [Gender]- Student addressed gender diversity of individuals within the course or of the instructor team, or how the instructor or other students approached gender diversity. Ex. “As a female I felt like the minority a lot of the time which feeds into my feelings of incompetence in the course.”

Demographics/Background [Sexual Orientation]- Student addressed sexual orientation of individuals within the course or of the instructor team, or how the instructor team or other students approached sexual orientation. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [Race/Ethnicity]- Student addressed racial/ethnic diversity of individuals within the course or of the instructor team, or how the instructor team or other students approached racial/ethnic diversity. Ex. “I never felt excluded in class because of my ethnicity”

Demographics/Background [Religion]- Student addressed religious diversity of individuals within the course or of the instructor team, or how the instructor team or other students approached religious diversity. Ex. “I did not feel like my religious observance was respected. When I asked about the possibility of taking a test earlier because it conflicted with my religious practice I was scoffed at and felt like a part of life very important to me was being ignored and disrespected. I did not need to be allowed to take the test at a different time, but the response I received discounted my faith and the teacher assumed she could understand my religion and the way I practice which was unfair and hurtful”

Demographics/Background [Politics]- Student addressed political ideologies of individuals within the course or of the instructor team, or how the instructor team or other students approached political ideologies. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [Year]- Student addressed the academic year of individuals within the course or how the instructor team or other students approached students’ academic years within the course. Ex. “I felt like there were a wide range of students from multiple classes (freshman, sophomores, juniors, etc.) and from multiple areas of study.”

Demographics/Background [Academic Background]- Student addressed prior academic background of individuals within the course or of the instructor team, such as being a first-generation student, transfer student, type of high school they went to, etc. Student could also have addressed how the instructor team or other students approached prior academic background. Ex. “Dr. [] highlighting her own background as a public school student from Baltimore helped me feel more equipped to learn chemistry, having also come from a public school […]”

Demographics/Background [International]- Student addressed accommodations available for international students within the course. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [Non-traditional]- Student addressed accommodations available for non-traditional students within the course. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [Disability]- Student addressed accommodations available for those with different abilities that are not temporary within the course. Ex. “its very fast so i struggle (i have dyslexia)”

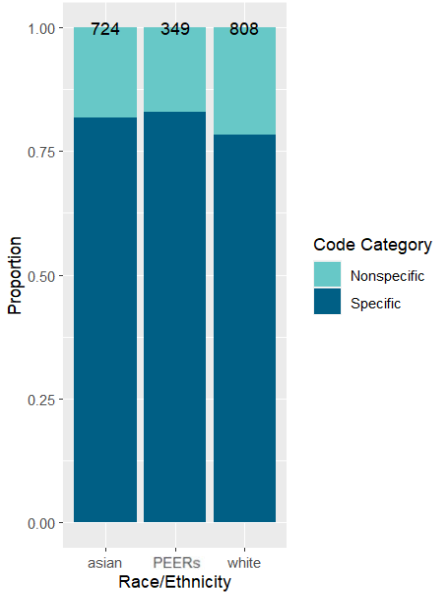
Demographics/Background [Dependents]- Student addressed accommodations available for those caring for dependents within the course. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [SES]- Student addressed accommodations available for those with different socioeconomic backgrounds within the course. Note: In this dataset, we found no quotes that fit this theme.

Demographics/Background [Majors]- Student addressed the experience within the course for students of various majors or STEM field. Ex. “[…]but as one of the many Chemical Engineers who for a reason I'm not entirely sure I understand are required to take this class, to say that I felt out of place is understatement [...]”

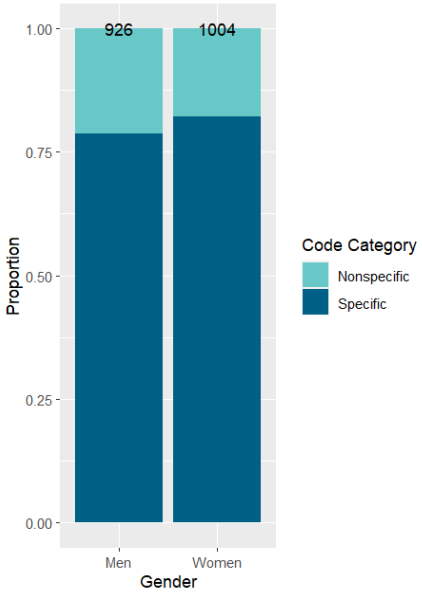
H. Supplemental Information from the Results of Study 1

Supplemental Figure 1: Specificity of Responses by Race/Ethnicity.



	Proportion		
	Asian	PEERs	White
Nonspecific	0.19	0.17	0.22
Specific	0.81	0.83	0.78

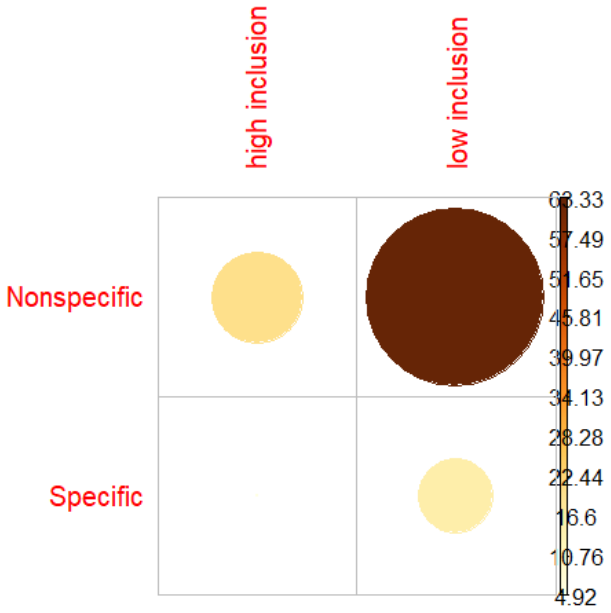
Supplemental Figure 2: Specificity of Responses by Gender.



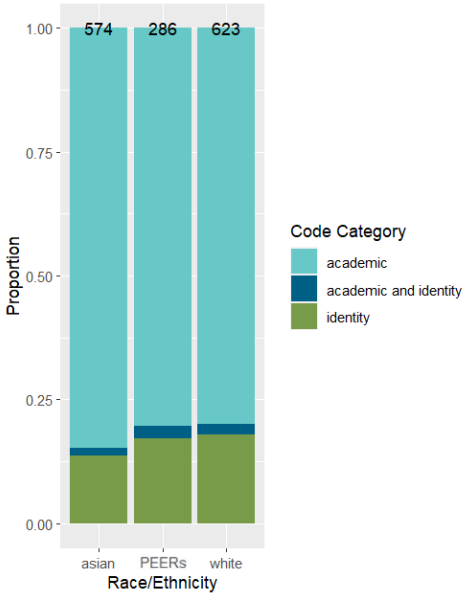
	Proportion	
	Men	Women
Nonspecific	0.22	0.18
Specific	0.78	0.82

Supplemental Figure 3:

Residual Figure Comparison for Chi-Squared test for specificity by low/high inclusion. With a Bonferroni correction, only nonspecific for low inclusion was significant (p= .001). Y-axis displays the chi-square value of each individual comparison. The blue range means that the group is more represented than expected, and the red range means that the group has lower values than expected. The darker and larger the circle is represents a more significant chi-square value.



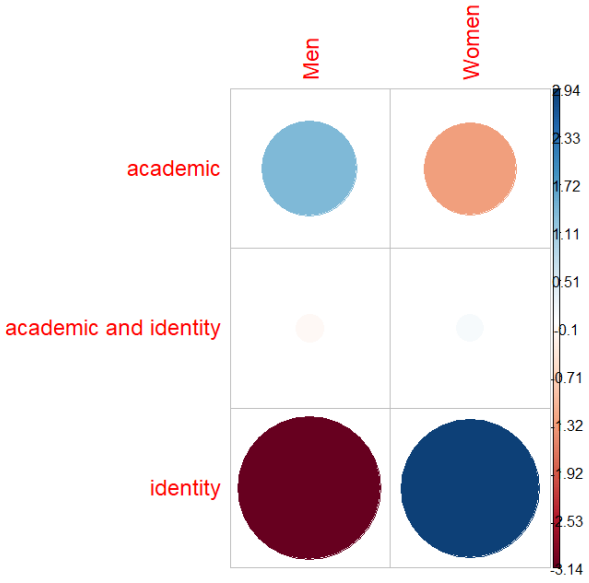
Supplemental Figure 4: Primary Categories Partitioned by Race/Ethnicity



	Proportion		
	Asian	PEERs	White
Academic	0.85	0.80	0.80
Academic and Identity	0.02	0.02	0.02
Identity	0.14	0.17	0.18

Supplemental Figure 5: Primary Categories Partitioned by Inclusion Score

	Proportion	
	High Inclusion	Low Inclusion
Academic	0.83	0.79
Academic and Identity	0.02	0.02
Identity	0.15	0.19

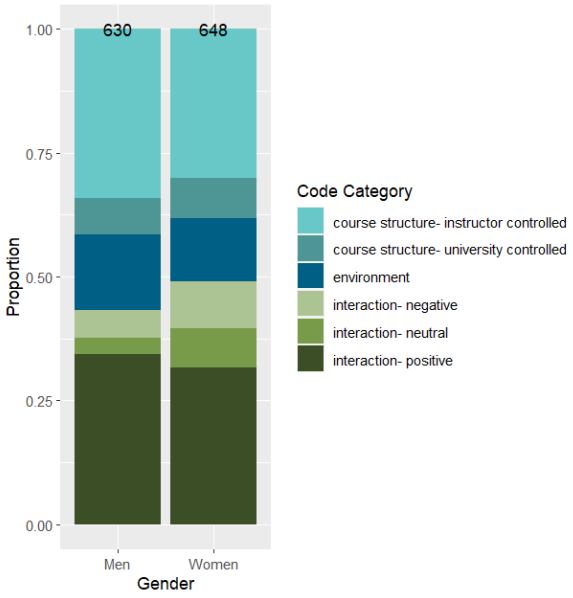


Supplemental Figure 6: Residual Figure Comparison for Chi-Squared test for primary categories by gender. With a Bonferroni correction, both identity by men and women were significant (p=0.002 and p=0.003, respectively). Y-axis displays the chi-square value of each individual comparison. The blue range means that the group is more represented than expected, and the red range means that the group has lower values than expected. The darker and larger the circle is represents a more significant chi-square value.

Academic Subcategories

Supplemental Figure 7: Breakdown of the Academic Subcategories across Gender. While we find that women seem to have a less positive experience with instructor interactions (10.2% versus 6.1% in males), this seems driven by an isolated incident in one classroom. Women also seem to cite student-instructor interactions more in their responses than men.

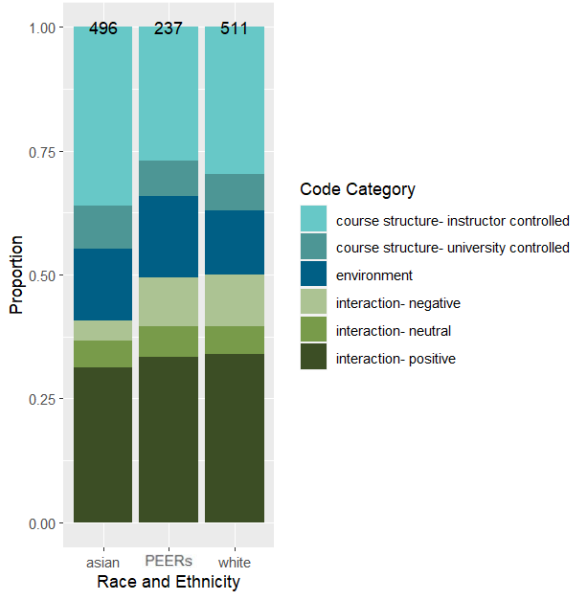
χ^2= 4.61, df = 2, p-value = 0.10



	Proportion	
	Men	Women
Environment	0.15	0.13
Student-Instructor Interaction - Negative	0.06	0.10
Student-Instructor Interaction - Neutral	0.03	0.08
Student-Instructor Interaction - Positive	0.34	0.32
Course Structure – Instructor-Controlled	0.34	0.30
Course Structure – University-Controlled	0.07	0.08

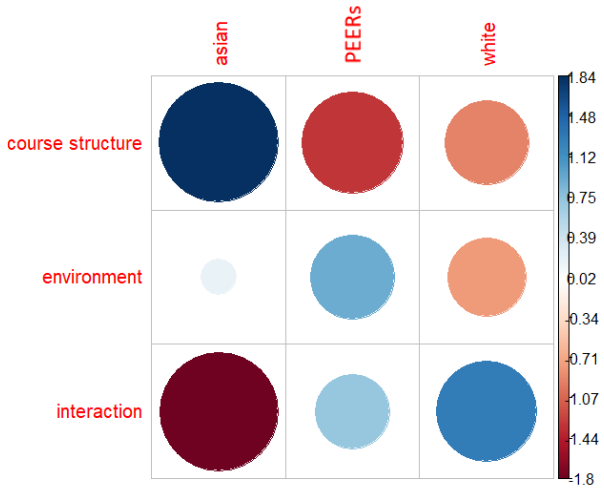
In comparing distribution of responses in subcategory by race and ethnicity, the overall pattern is significantly different, because multiple subcategories contribute to the differences (χ^2= 12.86, df = 4, p-value = 0.01; Cramer's V=0.072; Supplemental Figure 8). However, none of the individual cell comparisons reach significance after applying the Bonferroni correction (new threshold $p \leq 0.0056$).

Supplemental Figure 8: Breakdown of the Academic Subcategories by Race and Ethnicity



	Proportion		
	Asian	PEERs	White
Course Structure – Instructor Controlled	0.36	0.27	0.30
Course Structure – University Controlled	0.087	0.072	0.07
Environment	0.15	0.17	0.13
Student-Instructor Interaction - Negative	0.04	0.10	0.10
Student-Instructor Interaction - Neutral	0.05	0.06	0.06
Student-Instructor Interaction - Positive	0.31	0.33	0.34

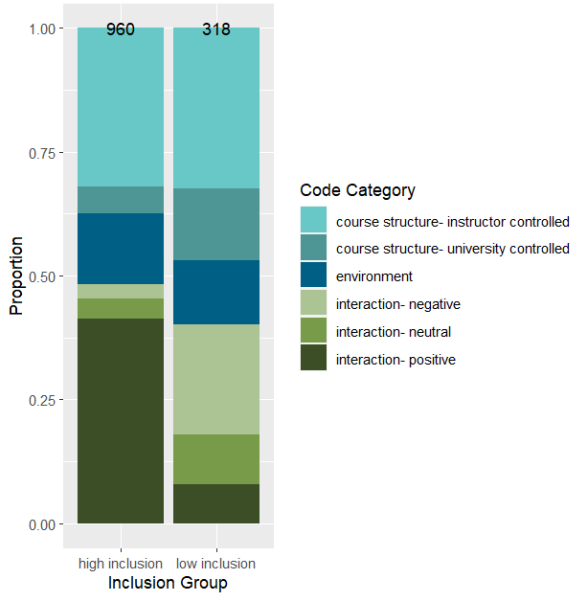
Supplemental Figure 9: Residual Plot for Chi-Squared Test for race/ethnicity. With a Bonferroni correction, none of the values are significant. Y-axis displays the chi-square value of each individual comparison. The blue range means that the group is more represented than expected, and the red range means that the group has lower values than expected. The darker and larger the circle is represents a more significant chi-square value.



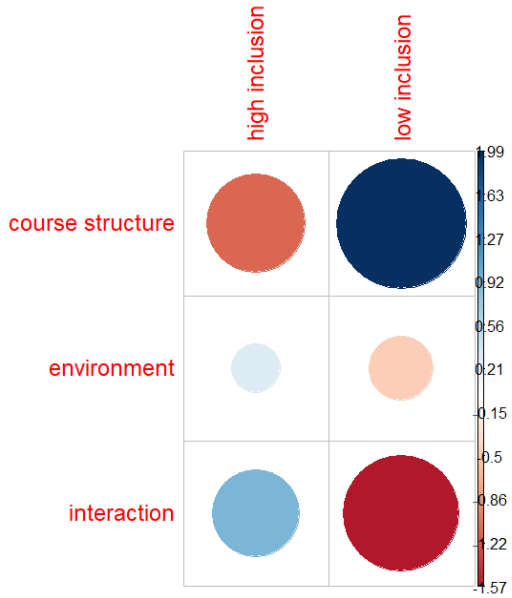
Course Inclusion Score

Similar to the observed subcategory response distributions by race and ethnicity above, comparisons by inclusion score also showed that these distributions differed significantly (Supplemental Figure 10). However, none of the individual subcategory differences rose to the level of significance after the Bonferroni correction was applied (new threshold $p \leq 0.0083$; Supplemental Part G, Supplemental Figure 11). Some differences observed are interesting despite not rising to statistical significance. Students who scored their courses lower on overall inclusion were more likely to cite course-structure elements that were not in an instructor's control, such as class size (15% versus 6% in the high inclusion-reporting group) and were less likely to refer to student-instructor interactions overall (40% versus 48%; Figure 10). Not unexpectedly, when students in the low inclusion-reporting group described 'student-instructor interaction' in their responses, they were more likely to describe neutral (10% versus 4%) or negative feelings (22% versus 3%) towards their interactions ("I feel like not everybody feels welcome to ask questions at times due to previous responses from the instructor").

Supplemental Figure 10: Breakdown of the Academic Subcategories by Low/High Overall Inclusion Score



	Proportion	
	High Inclusion	Low Inclusion
Course Structure – Instructor-Controlled	0.32	0.32
Course Structure - University Controlled	0.06	0.15
Environment	0.14	0.13
Student-Instructor Interaction - Negative	0.03	0.22
Student-Instructor Interaction - Neutral	0.04	0.10
Student-Instructor Interaction - Positive	0.41	0.08



Supplemental Figure 11: Residuals for Chi-Squared Tests for low/high inclusion. With a Bonferroni correction, none of the values are significant. Y-axis displays the chi-square value of each individual comparison. The blue range means that the group is more represented than expected, and the red range means that the group has lower values than expected. The darker and larger the circle is represents a more significant chi-square value.

Biology Only Data

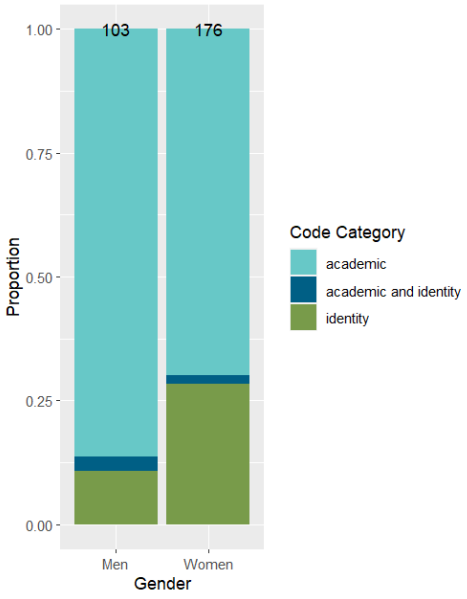
Supplemental Table 2: Biology Only versus Overall Data with the Primary Categories and Academic Subcategories. Although the chi-square test comparing primary categories across discipline (biology versus other STEM) is significant (χ^2= 8.94, df = 2, p-value = 0.011; Cramer's V=0.077), no individual cell frequency differed significantly from the expected frequency after a Bonferroni correction (new threshold $p \leq 0.0038$) was applied. The chi-square test comparing the academic subcategories across discipline was not significant (p=0.53).

	Proportion		
	Overall Data	Biology only	Other STEM only
Specific Responses	0.80	0.79	0.80
Nonspecific Responses	0.20	0.21	0.20
Academic	0.80	0.76	0.83
Academic and Identity	0.02	0.02	0.02
Identity	0.16	0.22	0.15

	Proportion		
	Overall	Biology only	Other STEM only
Course Structure – Instructor-Controlled	0.32	0.26	0.32
Course Structure – University-Controlled	0.08	0.08	0.08
Environment	0.14	0.13	0.14
Student-Instructor Interaction - Negative	0.08	0.22	0.05
Student-Instructor Interaction - Neutral	0.06	0.09	0.05
Student-Instructor Interaction - Positive	0.33	0.22	0.36

Small, significant differences across the primary categories were found across gender (χ^2= 12.116, df = 2, p= 0.002; Cramer's V= 0.208; Supplemental Figure 12), but not race/ethnicity (χ^2= 3.855, df = 4, p=0.426; Supplemental Figure 13) or overall low/high reported inclusion scores (χ^2= 4.788, df = 2, p=0.091; Supplemental Figure 14). After the Bonferroni correction was applied (new threshold $p \leq 0.0038$ for both gender and low/high inclusion-reporting groups), cell-wise post-hoc analyses confirmed that gender did not significantly influence the frequency of using any of the primary codes, but the proportion of men in the biology course who cited identity factors trended towards significance, with men being less likely to cite identity in their responses (p = 0.0048).

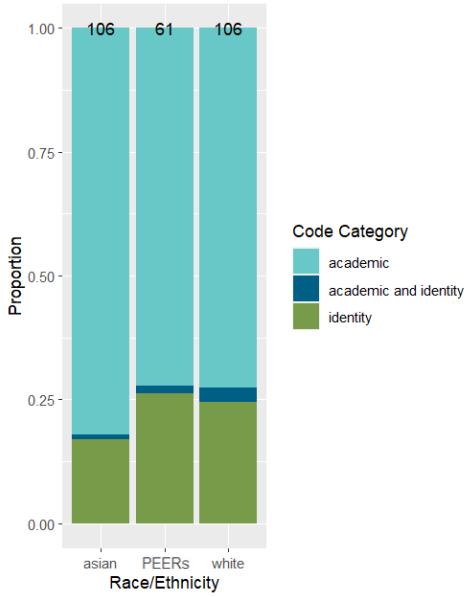
Supplemental Figure 12: Biology Course Primary Categories by Gender. Differences for gender (p= 0.002). Women were a lot more likely to discuss 'identity'. However, with a Bonferroni correction, none of the values are significant. χ^2= 12.116, df = 2, p= 0.002; Cramer's V= 0.208.



	Proportion	
	Men	Women
Academic	0.86	0.70
Acad and Identity	0.03	0.02
Identity	0.11	0.28

Supplemental Figure 13: Biology Course Primary Categories by Race/Ethnicity. This chi-square test is not significant (χ^2= 3.855, df = 4, p=0.426).

	Proportion		
	Asian	PEERs	White
Academic	0.82	0.72	0.73
Acad and Identity	0.01	0.02	0.03
Identity	0.17	0.26	0.25



Supplemental Figure 14: Biology Course Primary Categories by Low/High Inclusion. This chi-square test is not significant (χ^2= 4.788, df = 2, p=0.091).

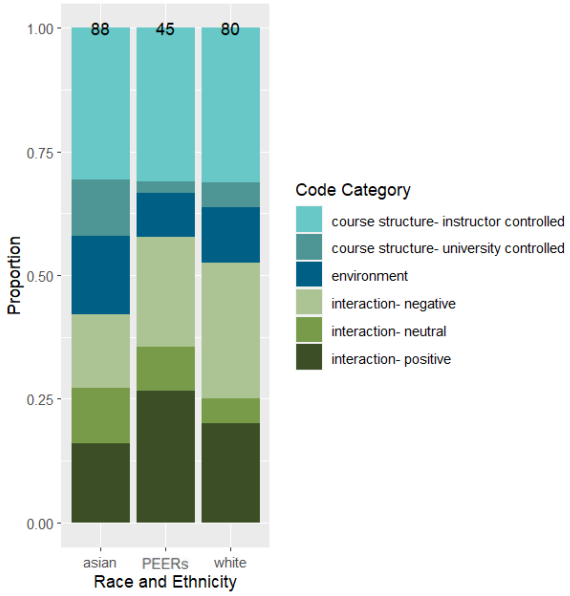
	Proportion	
	High Inclusion	Low Inclusion
Academic	0.73	0.80
Acad and Identity	0.01	0.03
Identity	0.26	0.16

Academic Subcategories in the Biology Course

Across the academic subcategories in the introductory biology course data set, the chi-square tests between the different subgroups (gender, race/ethnicity, course-inclusion level) and the three subcategories yielded no significant differences. Despite the difference not rising to the level of significance, we felt some differences should be noted. The pattern for race/ethnicity is similar between biology and other STEM disciplines, but not identical —biology PEERs students were more likely to cite 'student-instructor interaction' (65.8% versus 53.2% in the overall dataset) and less likely to cite factors that fell into the 'environment' subcode than in other STEM courses (Supplemental Figure 15). The pattern across gender in biology mirrors the other STEM courses (and the overall data set). For low/high inclusion-reporting groups, students who reported high inclusion scores in biology were descriptively more likely to talk about factors outside an instructor's control (i.e., primarily class size) than students who reported low inclusion scores in biology (Supplemental Figure 16).

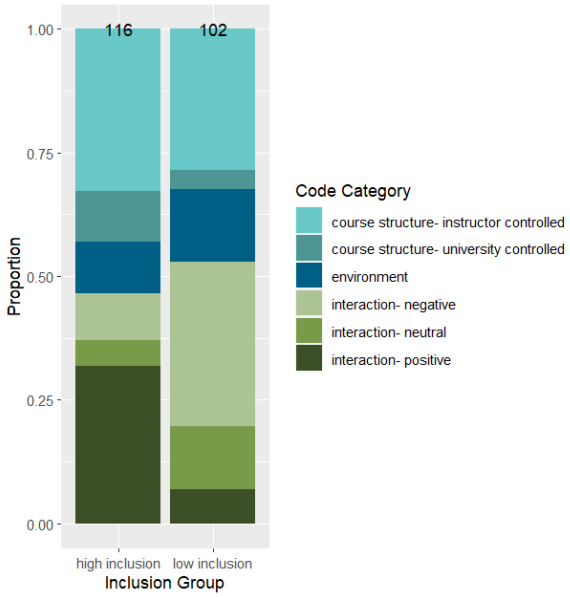
Supplemental Figure 15: Biology Course Academic Subcategories by Race/Ethnicity.

Descriptively, pattern for race is similar, but not the same. Asian students still more likely to describe university-controlled course structure factors , but PEER students no longer seem to discuss environment more. PEERs are more likely to describe student-instructor interaction than whites, which are more likely to discuss student-instructor interaction than Asian students.



	Proportion		
	Asian	PEERs	White
Environment	0.16	0.09	0.11
Course Structure – University-Controlled	0.11	0.00	0.05
Course Structure – Instructor-Controlled	0.31	0.32	0.31
Student-Instructor Interaction - Negative	0.15	0.23	0.26
Student-Instructor Interaction - Neutral	0.11	0.09	0.06
Student-Instructor Interaction - Positive	0.16	0.27	0.20

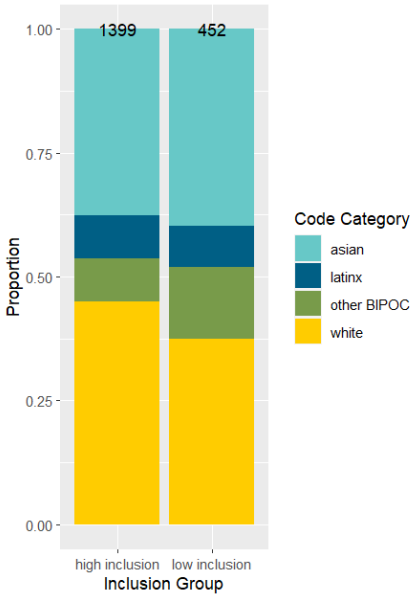
Supplemental Figure 16: Biology Course Academic Subcategories by Inclusion Group. Students who score inclusion lower are still slightly more likely to describe student-instructor interactions. However, we find that students who score inclusion higher are more likely to talk about factors outside of the instructor's control than those who score overall inclusion lower.



	Proportion	
	High Inclusion	Low Inclusion
Environment	0.10	0.15
Course Structure – University-Controlled	0.10	0.04
Course Structure – Instructor-Controlled	0.33	0.28
Student-Instructor Interaction - Negative	0.10	0.32
Student-Instructor Interaction - Neutral	0.05	0.14
Student-Instructor Interaction - Positive	0.32	0.07

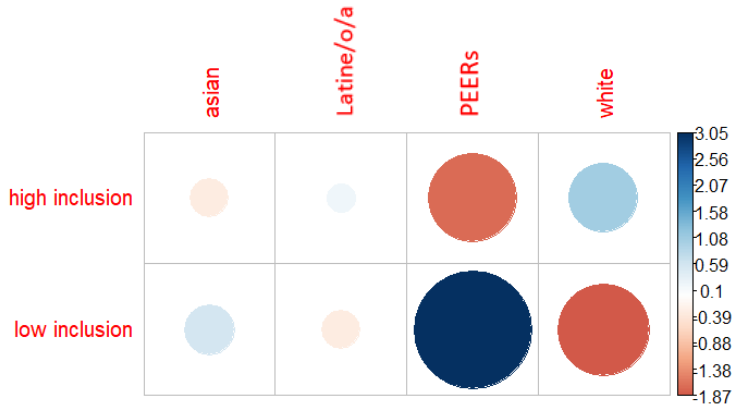
χ^2= 2.928, df = 2, p=0.231

Supplemental Figure 17: Inclusion Score Group Breakdown by Race (4 Groups). χ^2= 17.55, df = 3, p=0.00055.



Race/Ethnicity	Proportion	
	High Inclusion Students	Low Inclusion Students
Asian	0.38	0.34
Latine/o/a	0.09	0.08
Other PEERs	0.09	0.15
White	0.44	0.37

Supplemental Figure 18: Residual Plot for Chi-Squared Test for Inclusion Score by Race (Four groups). With a Bonferroni correction, there is a significant result for more students who identify as 'other PEERs' to rate their experience in the course lower in overall inclusion score. Y-axis displays the chi-square value of each individual comparison. The blue range means that the group is more represented than expected, and the red range means that the group has lower values than expected. The darker and larger the circle is represents a more significant chi-square value.



Supplemental Table 3: Average Exam Performance Effects by Course in Overall Dataset.

Course	High Inclusion Mean	Low Inclusion Mean	t	df	p	Effect size
Introductory Biology 1	83.27	79.98	2.80	268.63	0.00	0.31, small
General Chemistry 1	75.37	71.55	1.93	90.77	0.06	0.29, small
General Chemistry 2	77.17	75.87	0.62	90.87	0.54	0.09, negligible
Introductory Physics 1	83.04	77.32	4.72	205.47	0.00	0.50, medium
Introductory Physics 2	82.52	78.95	2.22	86.24	0.03	0.32, small

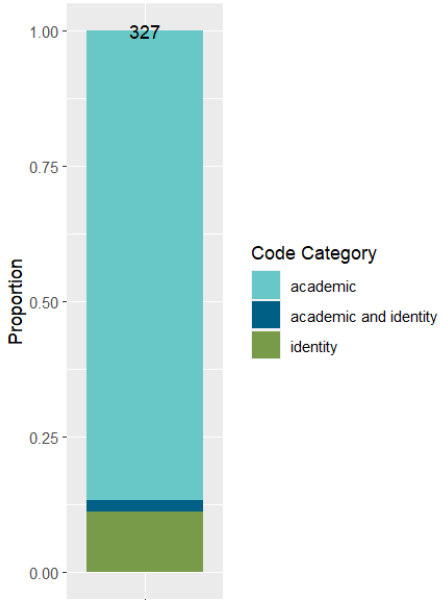
Supplemental Table 4: Average Exam Performance Effects by Course in Academic Category

Course	High Inclusion Mean	Low Inclusion Mean	t	df	p	Effect size
IB1	83.46	80.54	1.98	186.73	0.05	0.28, small
GC1	74.42	70.64	1.54	60.02	0.13	0.29, small
GC2	78.12	79.33	-0.51	63.27	0.61	-0.10, negligible
IP1	82.64	75.28	4.44	108.30	0.00	0.63, medium
IP2	82.27	77.55	2.25	48.96	0.03	0.42, small

I. Supplemental Information from the Results of Study 2

Supplemental Figure 19: Overall Breakdown of Primary Categories in Cohort of Study 2 Students versus Whole Dataset.

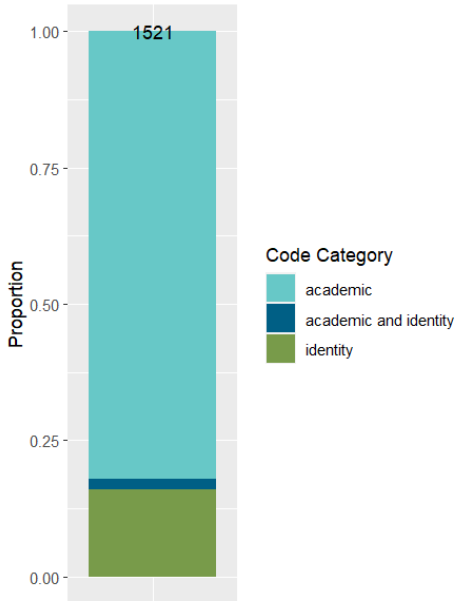
Overall Plot- Subset



	Proportion
Academic	0.87
Acad and Identity	0.02
Identity	0.11

χ^2= 7.66, df = 2, p=0.02

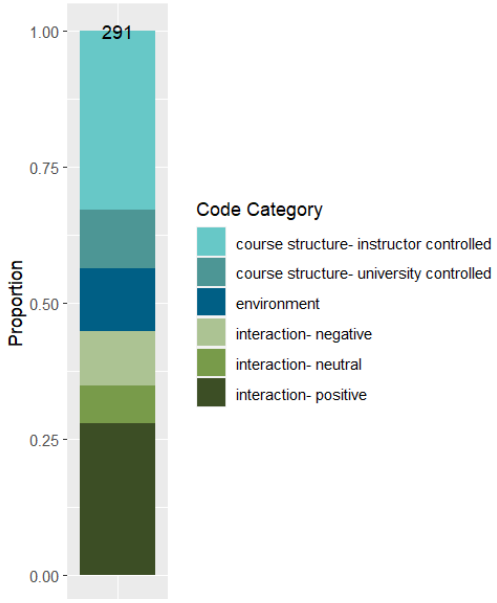
Overall Data- Whole Set



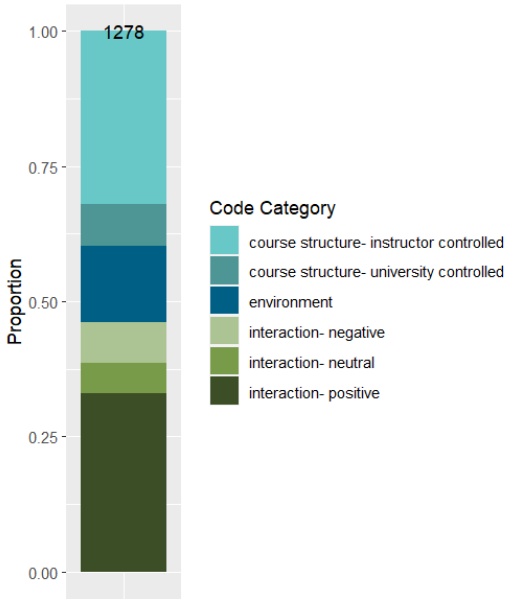
	Proportion
Academic	0.82
Acad and Identity	0.02
Identity	0.16

Supplemental Figure 20: Overall Breakdown of Academic Subcategories in Cohort of Study 2 Students versus Whole Dataset.

Academic Subcategory- Subset



Academic Subcategory- Whole Set

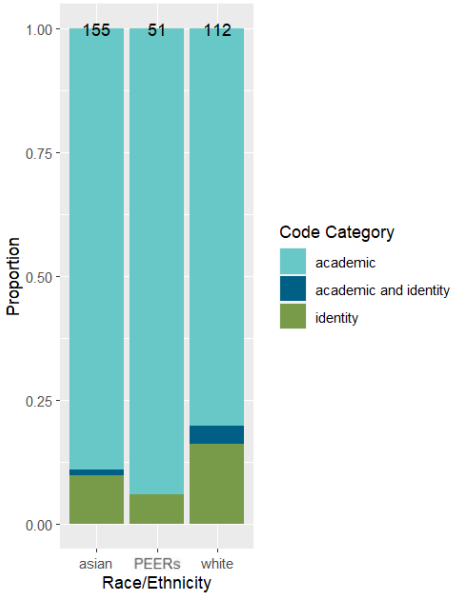


χ^2= 0.99, df = 2, p=0.61

Subset	Proportion
Course Structure – Instructor-Controlled	0.32
Course Structure – University-Controlled	0.10
Environment	0.13
Student-Instructor Interaction - Negative	0.11
Student-Instructor Interaction - Neutral	0.07
Student-Instructor Interaction - Positive	0.28

Overall	Proportion
Course Structure – Instructor-Controlled	0.321
Course Structure – University-Controlled	0.08
Environment	0.14
Student-Instructor Interaction - Negative	0.08
Student-Instructor Interaction - Neutral	0.06
Student-Instructor Interaction - Positive	0.33

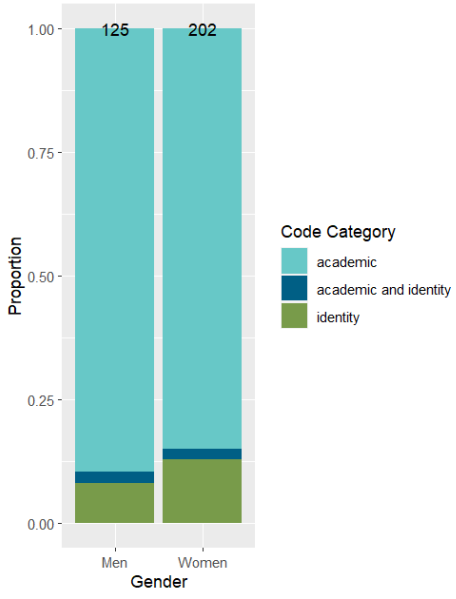
Supplemental Figure 21: Study 2 Cohort Breakdown by Race/Ethnicity in Primary Categories



	Proportion		
	Asian	PEERs	White
Academic	0.89	0.94	0.80
Acad and Identity	0.01	0.00	0.04
Identity	0.10	0.06	0.16

χ^2= 7.81, df = 4, p-value = 0.10

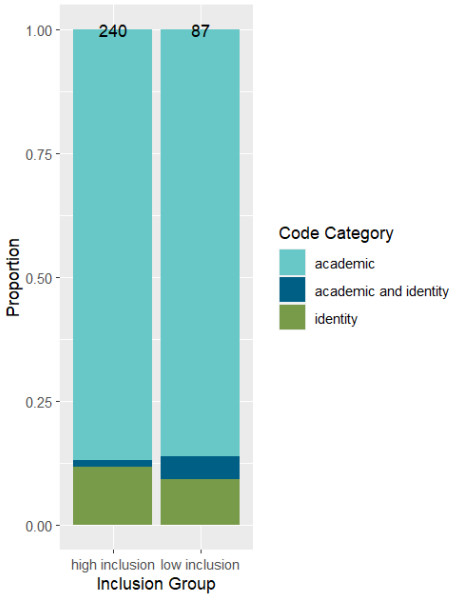
Supplemental Figure 22: Study 2 Subset Breakdown by Gender in Primary Categories



	Proportion	
	Men	Women
Academic	0.90	0.85
Acad and Identity	0.02	0.02
Identity	0.08	0.13

χ^2= 1.90, df = 2, p-value = 0.39

Supplemental Figure 23: Study 2 Subset Breakdown by Students with Low/High Inclusion scores in Primary Categories.



	Proportion	
	High Inclusion	Low Inclusion
Academic	0.87	0.86
Acad and Identity	0.01	0.05
Identity	0.12	0.09

χ^2= 3.70, df = 2, p-value = 0.157

J. Full Survey

Inclusion Survey for Students

While completing this survey, please think specifically about the lecture component of [insert course name], rather than laboratory, recitation, or any supplemental programs. The instructors of the course will not see your survey responses; they will only see a summary of the results after grades are submitted. You may skip any questions you do not wish to answer.

	Not at all	Slightly	Somewhat	Moderately	Highly
1. How inclusive do you feel the course [insert course name] was overall?					

2. Please explain your reasoning for your response to the previous question (2-3 sentences).

Please indicate the degree to which you agree that the instructors of [insert course name] do each of the following when teaching the course:	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
3. Create a classroom environment that is conducive to student participation.						
4. During class, use activities that encourage all students to participate.						
5. Evaluate student learning using multiple types of assessments (e.g., homework, quizzes, tests, presentations).						
6. Include details about course policies, course requirements, course schedule and course deadlines in the syllabus or general course information.						
7. Remind students of upcoming course deadlines and events (e.g., upcoming exam and assignment due dates, help sessions and office hours).						
8. Try to ensure that all students feel a sense of belonging in the classroom.						
9. Present examples, resources, images, etc. that reflect a diverse population.						

10. Communicate to students the course philosophy and expectations.						
11. Design assignments that have clear instructions.						
12. Convey the idea that all students can learn and improve.						
13. Create an atmosphere of respect for all students.						
14. During class, give instructions about appropriate group interactions.						
15. During class, give instructions to promote equal participation within groups.						
16. Provide useful feedback on student assessments.						
17. Use inclusive language (e.g., do not always use the pronoun he).						
18. During class, establish explicit ground rules for appropriate classroom conduct.						
19. Teach in ways that do not reinforce negative stereotypes.						

20. In [insert course name], please estimate the percentage of class periods during which you individually asked or answered a question out loud. [students will use a slider with range 0 – 100% to respond to this question.]

21. Did any inappropriate behaviors or remarks ever occur during class? Yes or No If yes, answer question below. If no, skip to final question						
	Strongly Disagree	Somewhat Disagree	Slightly Disagree	Slightly Agree	Somewhat Agree	Strongly Agree
21a. The inappropriate behaviors or remarks were addressed.						
21b. Please explain your answers to the previous questions 21 and 21a, if you feel comfortable.						

22. Do you have any additional comments or feedback about the inclusivity of [insert course name]?

Bibliography

Frey, R. F., Fink, A., Cahill, M. J., McDaniel, M. A., & Solomon, E. D. (2018). Peer-Led Team Learning in General Chemistry I: Interactions with Identity, Academic Preparation, and a Course-Based Intervention. *Journal of Chemical Education*, *95*(12), 2103–2113. doi: 10.1021/acs.jchemed.8b00375

Frey, R. F., & Lewis, S. E. (2023). Implementing Cooperative Learning in Large Undergraduate Classes via PLTL. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars want you to know about why and how to apply the science of learning in your academic setting.* (pp. 239–251). Society for the Teaching of Psychology.

Hockings, S. C., DeAngelis, K. J., & Frey, R. F. (2008). Peer-led team learning in general chemistry: Implementation and evaluation. *Journal of Chemical Education*, *85*(7), 990–996. doi: 10.1021/ed085p990

Shields, S. P., Hogrebe, M. C., Spees, W. M., Handlin, L. B., Noelken, G. P., Riley, J. M., & Frey, R. F. (2012). A transition program for underprepared students in general chemistry: Diagnosis, implementation, and evaluation. *Journal of Chemical Education*, *89*(8), 995–1000. doi: 10.1021/ED100410J/SUPPL_FILE/ED100410J_SI_001.PDF

York, A. M., Fink, A., Stoen, S. M., Walck-Shannon, E. M., Wally, C. M., Luo, J., … Frey, R. F. (2021). Gender inequity in individual participation within physics and science, technology, engineering, and math courses. *Physical Review Physics Education Research*, *17*(2), 020140. doi: 10.1103/PHYSREVPHYSEDUCRES.17.020140/FIGURES/6/MEDIUM