

## **Student belonging in STEM courses that use group work**

**Paul J. Emigh**

*Department of Physics, Oregon State University, 301 Weniger Hall, Corvallis, Oregon, 97331-6507*

**Sujata Krishna**

*Department of Physics, University of Florida, 2001 Museum Road, Gainesville, Florida, 32611*

**Jiehong Liao**

*Bioengineering Program, Florida Gulf Coast University,  
10501 FGCU Boulevard South, Fort Myers, Florida, 33965*

**Katsuhiko Kita**

*Department of Biology, St. Francis College, 179 Livingston Street, Brooklyn, New York, 11201*

**Jennifer R. Casey**

*Department of Chemistry and Biochemistry, University of California,  
Los Angeles, 607 Charles E. Young Drive East, Los Angeles, California, 90095*

**Jayson Nissen**

*Nissen Education Research and Design, Slidell, Louisiana, 70458*

Students' sense of belonging predicts their success and persistence in STEM courses. Collaborative, small-group activities form the foundation of many research-based instructional strategies. Our broader project seeks to understand the role of small groups in students' sense of belonging to support instructors in the formation of equitable groups in active engagement classrooms. In this article, we focus on the construct and discrimination validity of a belonging measure. To assess the belonging measure's ability to discriminate across time, courses, and demographic groups, we administered a short survey on belonging in a variety of STEM courses that used groups as a pre- and post-class assessment. We analyzed the results using structural equation modeling to inform the validity of the survey and identify possible differences of interest. The results provided evidence for both construct and discrimination validity. Belonging varied across the courses and changed from pre to post in two of the four courses: one course saw a decrease and the other course saw an increase. Men tended to have a higher sense of belonging than women and the changes in belonging increased these gender differences. One possibility is that the differences observed across courses could result from the different practices used to support group work within each course. The validity evidence for the belonging measure indicates it will support our ongoing research to establish the statistical relationships between instructor practices to implement and support small groups and students' sense of belonging.

## I. INTRODUCTION

Student belonging in the classroom affects persistence in STEM, especially for women and underrepresented minorities [1, 2]. In particular, past research links student belonging to higher self-efficacy and motivation [3, 4], more engagement [5], and increased persistence [6–9]. Since not all students feel like they belong in STEM [10–14], some instructors have incorporated interventions designed to support student belonging, often to good effect [15, 16]. Many such interventions use student groups to leverage the connection between peer groups and sense of belonging [8, 14, 17, 18].

Student-centered, collaborative instruction can lead to more learning and higher grades than lecture-based instruction [19–22]. Such instruction may also achieve improved equity [23]. These practices often have students work in groups of three to five students. Best practices for small group work are not well established. For example, some studies support the formation of groups with heterogeneous prior achievement [23–25], others argue for creating groups with homogeneous prior achievement [26–29], and a few suggest it does not matter [30, 31]. Another common grouping strategy is the avoidance of solo status for students with underrepresented identities [32]. Laboratory research and stereotype threat [33–35] both indicate the harm that solo status can cause. We are not, however, aware of any studies in the classroom that indicate this harm occurs or the size of the harm.

Our overall project investigates the effectiveness of various grouping strategies instructors can use to form intentional groups that contribute to student outcomes and classroom equity. In this article we focus on measuring students’ sense of belonging in their class. While our broader project investigates a variety of student outcomes including self-efficacy, conceptual understanding, and course grade, this article focuses on *belonging*. Belonging is correlated with critical metrics like students’ persistence in STEM [1, 2]. Specifically, this article focuses on substantiating the construct and discrimination validity of a belonging measure [10] in a variety of STEM courses. Our specific research questions are:

1. To what extent did the measures of belonging show evidence for construct validity?
2. Did the belonging measures differ across courses in four STEM disciplines before and after instruction?
3. Did the belonging measures differ across gender before and after instruction?

## II. THEORETICAL FRAMEWORK

We use Connell & Wellborn’s self-system model of motivational development, which they based on the fundamental human needs for competence, autonomy, and belonging [36]. This model supposes that the classroom impacts students’ self-perception, including their sense of belonging, which in turn impacts their engagement and achievement [3, 36–38]. We further draw on an anti-deficit approach with the perspec-

tive that students enter a course with abilities, talents, and capital, and that the institution is then responsible for structuring a course to support all students [39–42].

## III. METHODS

We studied instructional practice and students’ sense of belonging in four courses. In each course, instructors randomly assigned groups of 3–5 students. All courses used frequent active learning in the classroom and encouraged student-student interactions around content. We surveyed students during the first and last 2 weeks of the course. Table I summarizes group work practices in each course, including the use of undergraduate learning assistants (LAs), who support instructors by co-facilitating small group work [44]. The use of LAs in the classroom has been shown to strengthen students’ sense of belonging in STEM [45, 46].

Table II gives the course demographics for students who consented to participate in the study. Data collection included many social identity groups, which we reduced in the table due to small sample sizes. Underrepresented minorities included Black, Hispanic, Indigenous, Pacific Islander, and Middle Eastern students. Gender included responses from non-binary and trans students, but we were unable to perform statistical analysis for these groups due to a small sample size.

Course A was an algebra-based physics I course with 690 health-related students at a large 4-year public university, mainly taken by students in their final two years. Each week the course met for two 50-minute lectures—including conceptual questions, worked examples, demonstrations, and think-pair-share iClicker questions—and two discussion section meetings with a TA—including homework help, quizzes, and group work. Students received participation points for working together (during about 25% of contact time) in assigned groups of 5 (with 6 groups per TA) on worksheets solving problems.

Course B was a core, upper-level biology course for majors with 62 students at a small, private, 4-year liberal-arts college. The college is a Hispanic-serving institution. The course consisted of three 55-minute lectures and a three-hour lab each week. The instructor lectured with frequent short discussions, with about 20% of contact time spent on group work, and there was an out of class group assignment.

TABLE I. Instructional practices in each course [43].

Course	A	B	C	D
Learning Assistants (LAs) support group work	-	-	X	X
Groups discuss expectations and draft contract	-	X	-	-
Groups work on structured assignments in class	X	X	X	X
Group members assigned changing roles	-	-	-	X
Groups work together outside of class	-	X	X	X
Peer evaluation	-	-	X	-

TABLE II. Demographic distribution by course for participants.

Course	Time	Asian	URM	White	Asian	URM	White
		Women	Women	Women	Men	Men	Men
A	Pre	56	121	166	30	46	65
A	Post	48	96	130	22	34	48
B	Pre	5	27	5	2	9	2
B	Post	5	26	4	2	9	2
C	Pre	3	26	19	4	36	40
C	Post	2	17	16	3	25	29
D	Pre	104	65	58	44	16	19
D	Post	77	40	45	33	11	14

Course C was a sophomore-level engineering course with sections of about 30 students each (177 total students), at a mid-sized, 4-year, public, emerging Hispanic-serving institution. Classes met three times a week for 100 minutes. Instructors presented the topic through an interactive lesson on the whiteboard with example problems. Students used 50% of contact time to work on problems at their tables in groups of about 4 as the instructor and one to two LAs circulated to facilitate group work and engage students in discussion regarding their problem-solving. There were about 3-4 groups per LA. In addition to these ungraded group activities, students worked on a graded, out-of-class project.

Course D was the first course in the general chemistry series for life science majors at a 4-year, research-intensive, public university. The course was taught by an instructor, four TAs, and approximately 20 LAs, with around 130 students enrolled in a section (266 total students). The course consisted of three 50-minute lectures and a 110-minute discussion each week during a ten-week quarter. The lecture used think-pair-share and clickers. The discussion section was based on a blended model of Process Oriented Guided Inquiry (POGIL) and Peer-Led Team Learning (PLTL). Students worked in small groups on structured worksheets designed around the learning cycle. LAs supported about two groups each. Teams also completed the second stage of the two midterms together.

We used a six-item instrument to measure social belonging on a six-point Likert scale [10] (strongly disagree, mostly disagree, somewhat disagree, somewhat agree, mostly agree, strongly agree). The instrument (see Table III) includes two factors: perceived belonging (first four items) and belonging uncertainty (last two items). Perceived belonging relates to the general feeling of belonging in a course, while belonging uncertainty relates to the stability of that belonging. We administered the pre- and post-course belonging surveys through the LASSO Platform for Courses A, B, and C; course D administered the surveys through their learning management system.

We used structural equation modeling (SEM) [47] to investigate the three research questions. For RQ1, we used confirmatory factor analysis (CFA) [47] to test the construct validity

of the belonging measures for our study population. We first tested the model based on Fink *et al.* [10] using the lavaan [48] package. When necessary to improve model fit, we used the modindices command. We repeated these steps until the factor loadings and the fit indices passed the cutoffs discussed in the next paragraph.

The root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and factor loadings informed how well the model fit the data. RMSEA is an absolute fit index that addresses parsimony in the model by accounting for the model’s degrees of freedom. High RMSEA indicates an over-constrained model with too few degrees of freedom. RMSEA has several proposed cutoffs,  $< 0.05$ ,  $0.06$ ,  $0.08$ , and  $0.10$ , indicating good to acceptable fit [49]. CFI and TLI are relative fit indices. They compare the test model to a baseline model with all covariances set to zero and all variances freely estimated. Both indices range from 0, no fit, to 1, best possible fit. We used a cutoff of CFI and TLI  $> 0.95$  [49]. Balancing absolute and relative fit indices can lead to a model that is simple and fits well. We report scaled fit indices. Hair *et al.* [50] proposes factor loadings of 0.5, explaining 25% of the variance in the item, as an absolute minimum and 0.7 as a preferred minimum, explaining 50% of the variance.

To test RQ2 and RQ3, we used multiple group SEM. This method built separate SEM models for each course (*i.e.*, multiple groups) to investigate the shift from pre- to post-instruction. To determine if differences existed across courses, we compared these SEMs to a set of SEMs that constrained the intercepts and regression coefficients to the same value across all of the courses. We used the RMSEA, CFI, and TLI to assess the model fit. To determine if the unconstrained multigroup SEMs provided unique information, we compared them to the constrained multigroup SEMs with an ANOVA. We set the intercept for the models to the largest group: pretest for RQ2 and women’s pretest for RQ3. The re-

TABLE III. Questions on the belonging survey with their factor loadings for the pretest and posttest.

Question	Factor Loading	
	Pre	Post
Perceived Belonging Questions		
I feel like I fit in this course.	0.9	0.9
Setting aside my performance, I feel like I belong in this course.	0.9	1.0
I feel comfortable with my peers and classmates in this course.	0.7	0.7
I feel comfortable with my instructor(s) in this course.	0.7	0.7
Belonging Uncertainty Questions		
I feel uncertain about my belonging in this course.	1.0	1.0
If I don’t perform, I feel like maybe I don’t belong in this course.	0.5	0.5

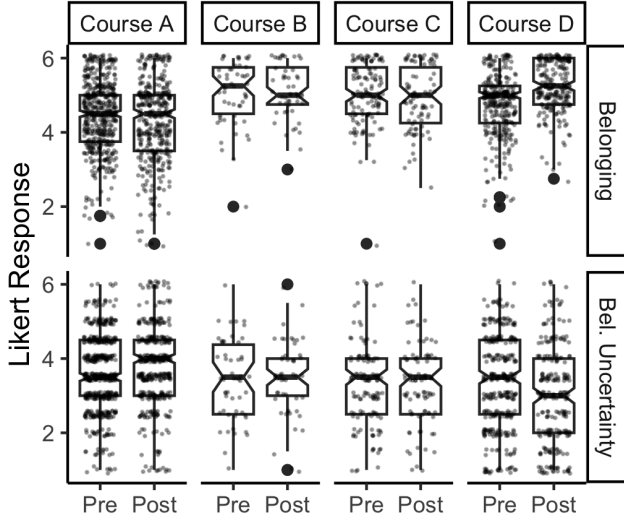


FIG. 1. Belonging across the four institutions constructed from the averages of the questions in the two factors, which are shown in Table III. The boxplots show the distribution of the data with outliers as large black dots. The light grey dots represent individual students with a slight jitter to spread out the data.

gression coefficients represent differences from the intercept in units of standard deviation (SD). The SEM builds the latent variables as normal continuous distributions. The SEM calculates thresholds for the point on this continuous distribution that a response would shift (e.g., going from strongly agree to agree) such that our six-response Likert-scale had five thresholds. The thresholds tended to cover a span of 3 SD, which means going from strongly disagree to strongly agree is approximately a 3 SD shift. Because of this scale and the spread of the results in the data, we adopted Cohen’s [51] rules of thumb for effect sizes:  $< 0.2$  is very small,  $0.2$  to  $0.4$  is small,  $0.4$  to  $0.8$  is moderate, and  $> 0.8$  is large.

#### IV. RESULTS

Figure 1 presents the perceived belonging and belonging uncertainty (pre and post) for the four courses constructed by averaging the questions for each factor. Perceived belonging tended towards ‘mostly agree’ in all courses. Perceived belonging was lower in Course A and highest on the posttest in Course D. The only notable shift in perceived belonging was for Course D where both the median and inter-quartile range increased. Belonging uncertainty tended to be between ‘slightly agree’ and ‘slightly disagree’. Consistent with perceived belonging, the noticeable shift in belonging uncertainty was for Course D where both the median and inter-quartile range shifted down, indicating higher belonging.

We tested the two-factor structure of the sense of belonging scale using CFA for the pretest and posttest data. All

of the fit indices indicated a good fit:  $CFI_{pre} = 0.997$ ,  $CFI_{post} = 0.978$ .  $TLI_{pre} = 0.982$ , and  $TLI_{post} = 0.953$  ( $> 0.950$  indicates good fit),  $RMSEA_{pre} = 0.039$  and  $RMSEA_{post} = 0.059$  ( $< 0.060$  indicates good fit). The factor loadings, shown in Table III, indicate adequate fits for all items with the lowest fit for the question about performance for the belonging uncertainty factor.

We investigated the change over time in the courses (RQ2) using multi-group SEM, and identified several relationships that also appear in Fig. 1 of the average scores. We did not include Course B in the analysis because no students responded to ‘strongly disagree’ or ‘disagree’ for several questions on belonging. The ANOVA between the constrained and unconstrained SEMs was statistically significant ( $p < 0.001$ ), indicating variation across the courses. The fit indices for the unconstrained SEM indicated adequate fit:  $TLI = 1.00$ ,  $CFI = 1.00$ ,  $RMSEA = 0.096$ . The regression coefficients in Table IV indicated small to moderate improvements in Course D, little to no change in Course C and very small decreases in belonging in Course A.

Due to small sample sizes, we restricted our analysis for RQ3 to investigate gender differences in Course A and D. The multigroup (course) SEM created standardized regression coefficients, shown in Fig. 2 and Table IV, for men’s pretest and posttest scores and women’s posttest scores. The intercept was women’s pretest scores. In both courses, men tended to report a higher belonging and lower belonging uncertainty than women both before and after instruction, in agreement with previous findings [11–13, 52, 53].

In course A, students’ perceived belonging decreased and belonging uncertainty increased. The decrease in perceived belonging for men shifted moderately (0.3 SD) and was statistically significant. The other shifts, while all being consistent with this decrease, were much smaller. The shifts in Course D were also consistent in that perceived belonging increased and belonging uncertainty decreased, and the shifts and dif-

TABLE IV. SEM regression coefficients for both models. The intercept was set to either the pretest or the pretest for women.

Course	Time	Gender	Belonging		Uncertainty	
			Coef.	SE	Coef.	SE
NA	Pre	-	0	NA	0	NA
A	Post	-	-0.181	0.072	0.122	0.069
C	Post	-	-0.149	0.154	0.003	0.114
D	Post	-	0.663	0.098	-0.314	0.097
NA	Pre	Women	0	NA	0	NA
A	Pre	Men	0.452	0.118	-0.316	0.101
A	Post	Women	-0.134	0.088	0.110	0.082
A	Post	Men	0.149	0.120	-0.269	0.108
D	Pre	Men	0.181	0.140	-0.323	0.147
D	Post	Women	0.574	0.119	-0.217	0.120
D	Post	Men	1.101	0.168	-0.217	0.120

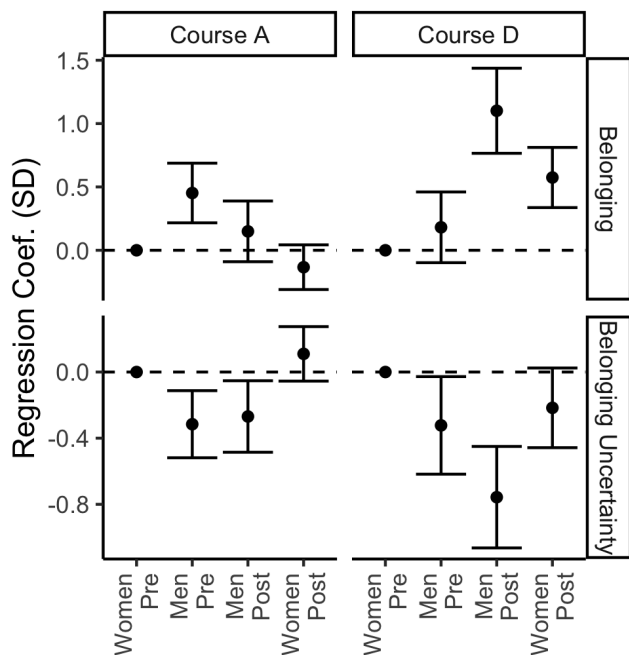


FIG. 2. Belonging for Course A and D for men and women. The scores are standardized within each course and compared to the pretest value for women which is constrained to zero. Error bars represent 2 SE. If the error bars for two coefficients do not contain the point estimates, the difference is likely statistically significant.

ferences across gender were much larger than in Course A. Women started (0.3 SD) and ended (0.5 SD) with higher belonging uncertainty than men and belonging uncertainty decreased for both men (0.4 SD) and women (0.2 SD). Women started (0.2 SD) and ended (0.5 SD) with lower perceived belonging than men and perceived belonging increased for both women (0.6 SD) and men (0.9 SD). Both the improvements in belonging and the greater improvements for men were consistent across both perceived belonging and belonging uncertainty. The shifts from pre to post were small to large in size. While belonging improved in Course D, these improvements were larger for the men in the course than for the women.

## V. DISCUSSION AND CONCLUSIONS

The CFA provided evidence for the construct validity of the belonging survey. All of the indicators of model fit met or exceeded the minimum values. The belonging measure also captured differences across courses, shifts in belonging for groups of students, and differences across demographic groups. The two latent factors, perceived belonging and belonging uncertainty, had very different average responses.

Perceived belonging was mostly positive across courses, with students tending to ‘mostly agree’ that they belonged in the course. Responses to belonging uncertainty tended towards ‘slightly disagree’ with a larger spread in all of the courses than the perceived belonging responses. The changes also varied across the courses, but were consistent for both belonging and belonging uncertainty. In courses B and C, there was little shift in either perceived belonging or belonging uncertainty. A small negative shift occurred in course A. The largest increases occurred in Course D with an increase in perceived belonging and a decrease in belonging uncertainty.

The most noteworthy shifts in student belonging occurred in course D, in which perceived belonging increased and belonging uncertainty decreased for both men and women. On the other hand, course A experienced shifts in the opposite direction, though they were not as large. We suspect the use of instructional practices oriented toward supporting group work (shown in Fig. 1) may contribute to the students’ sense of belonging in these courses, especially given that course A made only limited use of structured group assignments in class. Prior research suggests that one of these practices, the use of LAs, may be particularly effective at improving students’ sense of belonging [45, 46]. However, course C experienced at most small shifts, despite use of LAs. Because there are many other differences between the courses, including institution and field of study, we hope that expanding the courses in our dataset will help identify the factors contributing to belonging.

The gender differences present in the chemistry course (D) and not the physics course (A) stand out from prior work showing larger gender differences on many affective outcomes in physics than in chemistry [54, 55]. This finding also contrasts recent work showing that rotating roles can benefit women in particular [56, 57], as only Course D used this practice. While these results indicate discrimination validity in the belonging measure, the small sample in this study limits our ability to investigate the causes of these differences. The results do point to the need for the larger study across the intersections of discipline, gender, and race/ethnicity that we are pursuing.

These findings support our use of this belonging measure in our larger study of instructor practices to support group work. The construct and discrimination validity indicate the instrument can identify differences across courses and likely could identify differences between groups or students randomly assigned to different conditions. Ongoing data collection will support several investigations. Larger samples will allow modeling the relationships between the instructor practices in Table I and student belonging, which we can only speculate upon here based on prior research. Ongoing research will also look at differences in belonging and learning outcomes for groups. We will investigate if groups with either homogeneous or heterogeneous prior performance lead to different student outcomes and outcomes for students in solo-status conditions due to either their race or gender.



- 
- [1] V. Derricks and D. Sekaquaptewa, They're Comparing Me to Her: Social Comparison Perceptions Reduce Belonging and STEM Engagement Among Women With Token Status, *Psychology of Women Quarterly* **45**, 325 (2021).
- [2] S. Fischer, The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence, *Labour Economics* **46**, 211 (2017).
- [3] T. M. Freeman, L. H. Anderman, and J. M. Jensen, Sense of Belonging in College Freshmen at the Classroom and Campus Levels, *75*, 203 (2007).
- [4] 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, *Physical Review Physics Education Research* **12**, 020110 (2016).
- [5] D. Wilson, D. Jones, F. Bocell, J. Crawford, M. J. Kim, N. Veilleux, T. Floyd-Smith, R. Bates, and M. Plett, Belonging and Academic Engagement Among Undergraduate STEM Students: A Multi-institutional Study, *Research in Higher Education* **56**, 750 (2015).
- [6] C. Good, A. Rattan, and C. S. Dweck, Why do women opt out? Sense of belonging and women's representation in mathematics, *102*, 700 (2012).
- [7] L. R. M. Hausmann, J. W. Schofield, and R. L. Woods, Sense of belonging as a predictor of intentions to persist among african american and white first-year college students, *Research in Higher Education* **48**, 803 (2007).
- [8] S. Hurtado and D. F. Carter, Effects of college transition and perceptions of the campus racial climate on latino college students' sense of belonging, *Sociology of Education* **70**, 324 (1997).
- [9] M. Ong, C. Wright, L. Espinosa, and G. Orfield, Inside the Double Bind: A Synthesis of Empirical Research on Undergraduate and Graduate Women of Color in Science, Technology, Engineering, and Mathematics, *Harvard Educational Review* **81**, 172 (2011).
- [10] A. Fink, R. F. Frey, and E. D. Solomon, Belonging in general chemistry predicts first-year undergraduates' performance and attrition, *Chemistry Education Research and Practice* **21**, 1042 (2020).
- [11] D. Dortch and C. Patel, Black undergraduate women and their sense of belonging inSTEMat predominantly white institutions, *NASPA Journal About Women in Higher Education* **10**, 1 (2017).
- [12] K. L. Lewis, J. G. Stout, N. D. Finkelstein, S. J. Pollock, A. Miyake, G. L. Cohen, and T. A. Ito, Fitting in to Move Forward: Belonging, Gender, and Persistence in the Physical Sciences, Technology, Engineering, and Mathematics (pSTEM), *Psychology of Women Quarterly* **41**, 420 (2017).
- [13] 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, *International Journal of STEM Education* **5**, 10 (2018).
- [14] S. Rodriguez and J. Blaney, "We're the unicorns in stem: Understanding how academic and social experiences influence sense of belonging for latina undergraduate students., *Journal of Diversity in Higher Education* **14** (2020).
- [15] G. M. Walton, C. Logel, J. M. Peach, S. J. Spencer, and M. P. Zanna, Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering., *107*, 468 (2015).
- [16] G. M. Walton and G. L. Cohen, A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students, *Science* **331**, 1447 (2011).
- [17] A. W. Astin, What matters in college: Four critical years revisited. **22**, 10.2307/1176821 (1993).
- [18] S. Zumbrunn, C. McKim, E. Buhs, and L. R. Hawley, Support, belonging, motivation, and engagement in the college classroom: a mixed method study, *Instructional Science* **42**, 661 (2014).
- [19] 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, *Proceedings of the National Academy of Sciences* **111**, 8410 (2014).
- [20] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *66*, 64 (1998).
- [21] D. W. Johnson and R. T. Johnson, An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning, *Educational Researcher* **38**, 365 (2009).
- [22] L. Springer, M. E. Stanne, and S. S. Donovan, Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis, *69*, 21 (2016).
- [23] E. Theobald, M. Hill, E. Tran, S. Agrawal, N. Arroyo, S. Behling, N. Chambwe, D. Cintron, J. Cooper, G. Dunster, J. Grummer, K. Hennessey, J. Hsiao, N. Iranon, L. Jones, H. Jordt, M. Keller, M. Lacey, C. Littlefield, and S. Freeman, Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math, *Proceedings of the National Academy of Sciences* **117**, 201916903 (2020).
- [24] L. C. J. L. Brickell, L. C. D. B. Porter, L. C. M. F. Reynolds, and C. R. D. Cosgrove, Assigning Students to Groups for Engineering Design Projects: A Comparison of Five Methods, *Journal of Engineering Education* **83**, 259 (1994).
- [25] S. K. Horwitz and I. B. Horwitz, The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography, *Journal of Management* **33**, 987 (2007).
- [26] J. Baer, Grouping and Achievement in Cooperative Learning, *College Teaching* **51**, 169 (2003).
- [27] J. L. Jensen and A. Lawson, Effects of Collaborative Group Composition and Inquiry Instruction on Reasoning Gains and Achievement in Undergraduate Biology, *CBE - Life Sciences Education* **10**, 64 (2011).
- [28] S. B. Watson and J. E. Marshall, Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course, *Journal of Research in Science Teaching* **32**, 291 (1995).
- [29] N. M. Webb, K. M. Nemer, and S. Zuniga, Short circuits or superconductors? effects of group composition on high-achieving students' science assessment performance, *American Educational Research Journal* **39**, 943 (2002), <https://doi.org/10.3102/00028312039004943>.
- [30] J. J. Harlow, D. M. Harrison, and A. Meyertholen, Effective student teams for collaborative learning in an introductory uni-

- versity physics course, *Physical Review Physics Education Research* **12**, 010138 (2016).
- [31] F. Lawrenz and T. W. Munch, The effect of grouping of laboratory students on selected educational outcomes, *Journal of Research in Science Teaching* **21**, 699 (1984).
- [32] K. D. Tanner, Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity, *CBE - Life Sciences Education* **12**, 322 (2013).
- [33] K. I. Pollak and Y. F. Niemann, Black and White Tokens in Academia: A Difference of Chronic Versus Acute Distinctiveness1, *Journal of Applied Social Psychology* **28**, 954 (1998).
- [34] D. Sekaquaptewa, Solo Status: How Lack of Diversity Affects Members of Underrepresented Groups, in *ACS Symposium Series*, Vol. 1277, edited by R. Hernandez, D. Stallings, and S. K. Iyer (American Chemical Society, Washington, DC, 2018) pp. 141–158.
- [35] D. Sekaquaptewa and M. Thompson, Solo status, stereotype threat, and performance expectancies: Their effects on women’s performance, *Journal of Experimental Social Psychology* , 7 (2003).
- [36] J. P. Connell and J. G. Wellborn, Competence, autonomy and relatedness: a motivational analysis of self-system processes., in *Self Processes and Development*, edited by M. Gunnar and L. A. Sroufe (University of Chicago Press, Chicago, 1991) pp. 43–77.
- [37] M. Micari and P. Pazos, Connecting to the professor: impact of the student-faculty relationship in a highly challenging course, **60**, 41 (2012).
- [38] T. L. Strayhorn and M. C. Terrell, Mentoring and Satisfaction with College for Black Students, **58**, 69 (2007).
- [39] S. R. Harper, An anti-deficit achievement framework for research on students of color in STEM, *New Directions for Institutional Research* **2010**, 63 (2010), publisher: Wiley Online Library.
- [40] E. O. McGee, *Black, brown, bruised: How racialized STEM education stifles innovation* (Harvard Education Press, 2020).
- [41] J. Mejia, R. Revelo, I. Villanueva, and J. Mejia, Critical theoretical frameworks in engineering education: An anti-deficit and liberative approach, *Education Sciences* **8**, 158 (2018).
- [42] T. J. Yosso, Whose culture has capital? A critical race theory discussion of community cultural wealth, *Race Ethnicity and Education* **8**, 69 (2005).
- [43] Y. Chang and P. Brickman, When Group Work Doesn’t Work: Insights from Students, *CBE - Life Sciences Education* **17**, ar52 (2018).
- [44] V. Otero, S. Pollock, and N. Finkelstein, A physics department’s role in preparing physics teachers: The Colorado learning assistant model, *American Journal of Physics* **78**, 1218 (2010), publisher: American Association of Physics Teachers-SAAPT.
- [45] 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 theSTEMcommunity”: Incorporation of learning assistants enhances students’ sense of belonging in a large introductory biology course, *CBE-Life Sciences Education* **21**, ar26 (2022).
- [46] M. Wilton, E. Gonzalez-Niño, P. McPartlan, Z. Terner, R. E. Christoffersen, and J. H. Rothman, Improving academic performance, belonging, and retention through increasing structure of an introductory biology course, *CBE-Life Sciences Education* **18**, ar53 (2019).
- [47] K. A. Bollen, *Structural equations with latent variables*, Vol. 210 (John Wiley & Sons, 1989).
- [48] Y. Rosseel, lavaan: An R package for structural equation modeling, *Journal of Statistical Software* **48**, 1 (2012).
- [49] Y. Xia and Y. Yang, RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods, *Behavior research methods* **51**, 409 (2019).
- [50] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. Tatham, *Data Analysis Multivariate* (New Jersey. Upper Saddle River. Pearson Education, 2006).
- [51] J. Cohen, *Statistical power analysis for the behavioral sciences* (Academic press, 2013).
- [52] K. H. Almeida, Disaggregated general chemistry grades reveal differential success among bipoc students in partial flipped team learning classrooms, *Journal of Chemical Education* **99**, 259 (2021).
- [53] D. R. Johnson, Women of color in science, technology, engineering, and mathematics (stem), *New Directions for Institutional Research* **2011**, 75 (2011).
- [54] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, Why are someSTEMfields more gender balanced than others?, *Psychological bulletin* **143**, 1 (2017).
- [55] A. M. Kelly, Social cognitive perspective of gender disparities in undergraduate physics, *Physical Review Physics Education Research* **12**, 020116 (2016).
- [56] E. M. Stump, M. Dew, S. Jeon, and N. Holmes, Taking on a manager role can support women’s physics lab identity development, *Physical Review Physics Education Research* **19**, 010107 (2023).
- [57] N. Holmes, G. Heath, K. Hubenig, S. Jeon, Z. Y. Kalender, E. Stump, and E. C. Sayre, Evaluating the role of student preference in physics lab group equity, *Physical Review Physics Education Research* **18**, 010106 (2022).