Supplementary Information A.

Likert statements from the survey, organized by detected factor. *Items listed twice.

Factor	Item label	Survey item (Not at all descriptive, Minimally descriptive, Somewhat descriptive, Mostly descriptive, Very descriptive)
1	P1	I listen as the instructor guides me through major topics
1	P2	The class activities connect course content to my life and future work
1	P5	In my class a variety of means (models, drawings, graphs, symbols, simulations, tables, etc.) are used to represent course topics and/or solve problems
1	P11	Multiple approaches to solving a problem are discussed in class
1	P12	I have enough time during class to reflect about the processes I use to solve problems
1	P17	The instructor adjusts teaching based upon what the class understands and does not understand
1	P18	The instructor explains concepts in this class in a variety of ways
1	P19	I receive feedback from my instructor on homework, exams, quizzes, etc.
1	P22	My instructor uses strategies to encourage participation from a wide range of students
2	P6	I talk with other students about course topics during class
2	P7	I constructively criticize other students' ideas during class
2	P8	I discuss the difficulties I have with math with other students during class
2	P10	I work with other students in small groups during class
2	P15	Class is structured to encourage peer-to-peer support among students (e.g., ask peer before you ask instructor, having group roles, developing a group solution to share)
2	P16*	There is a sense of community among the students in my class
2	P20*	I share my ideas (or my group's ideas) during whole class discussions
3	P13	A wide range of students respond to the instructor's questions in class
3	P16*	There is a sense of community among the students in my class
3	P21	A wide range of students participate in class
4	P3	I receive immediate feedback on my work during class (e.g., student response systems such as clickers or voting systems; short quizzes)
4	P4	I am asked to respond to questions during class time
4	P20*	I share my ideas (or my group's ideas) during whole class discussions
	P9	I work on problems individually during class time
	P14	The instructor knows my name

Fall 2018

	Engagement	Collaboration	Equity
Collaboration	0.174 (SE*: 0.010)		
Equity	0.261 (SE: 0.012)	0.492 (SE: 0.019)	
Thinking	0.254 (SE: 0.013)	0.542 (SE: 0.023)	0.585 (SE: 0.022)

Spring 2019

	Engagement	Collaboration	Equity
Collaboration	0.242 (SE: 0.011)		
Equity	0.345 (SE: 0.013)	0.660 (SE: 0.019)	
Thinking	0.324 (SE: 0.013)	0.733 (SE: 0.022)	0.718 (SE: 0.021)

Fall 2019

	Engagement	Collaboration	Equity
Collaboration	0.144 (SE: 0.019)		
Equity	0.286 (SE: 0.026)	0.381 (SE: 0.035)	
Thinking	0.266 (SE: 0.029)	0.246 (SE: 0.035)	0.475 (SE: 0.039)

Spring 2020

	Engagement	Collaboration	Equity
	0.193 (SE: 0.020)		
Equity	0.298 (SE: 0.024)	0.407 (SE: 0.038)	
Thinking	0.242 (SE: 0.023)	0.340 (SE: 0.038)	0.489 (SE: 0.042)

*SE: standard error

Supplementary Information C: More Details on the Model Calculations

The following remarks provide additional details on a few of the decisions that were made in conducting the analysis of this paper that may be of interest to some readers.

As noted in the paper, the scree plot suggested a 3, 4, or 5 factor model. Ultimately a four factor model was chosen. Each of the 3F(actor), 4F, and 5F solutions produced a close fit. The RMSEA (Root Mean

Square Error of Approximation) values fell between 0.04 and 0.054; RMSEA values closer to 0 indicate closer fit. However, the 4F and 5F solutions both produced better fit statistics than those of the 3F solution. Though the 3F solution meets the Kaiser criterion, the use of the 3F model would use a factor supported solely on two variables and question groups unsupported by theory, as well as removal of one variable. Additionally, it has been suggested that the Kaiser criterion should be interpreted as a lower bound on the number of factors to use (Braeken & Assen 2016). The use of the 5F solution leads to less accurate results in confirmatory factor analysis (CFA), potentially due to overfitting. Using the 5F solution would also require the use of two factors using only two variables each.

In addition to the choice to use a four factor model, some readers may be interested in understanding the decision to allow cross loading in the model.

Cross loading traditionally creates problems with conventional metrics such as the use of average variance extracted (AVE) to determine discriminant validity were not suitable. However, two novel techniques have made it possible to still determine discriminant validity denoted by CI_{CFA}(cut) and $\chi^2{}_{CFA}$(cut).[1] (Rönkkö & Cho 2020). These two techniques allow for cross-loading in the model as long as assumptions are accounted for and the cross-loading is modeled correctly. Rönkkö and Cho suggest using CI_{CFA}(cut) over $\chi^2{}_{CFA}$(cut) to avoid misuse of the $\chi^2{}_{CFA}$(cut) technique and for simplicity of interpretation. Rönkkö and Cho also suggest using 0.85 as the cutoff for an upper bound on the confidence interval calculated for CI_{CFA}(cut) as other studies suggest that 0.85 offers the "best balance of high power and an acceptable Type I error rate" (Rönkkö & Cho 2020). Further description of the calculations can be found in (Rönkkö and Cho 2020).

[1] Available in the discriminantValidity function as part of the semTools package in R (Jorgensen et al. 2020).