**SUPPLEMENTARY MATERIALS**

**Item Factor Analysis**

To examine the psychometrics of our measures, we conducted an ordinal item factor analysis for each scale using the lavaan package in R. We have multiple reasons for choosing to use an IFA/classical test theory rather than IRT/Rasch analysis. First, studies have found that IRT and CTT have similar results in samples over 100 (e.g., Beaujean, 2014). The factor analysis we conducted can tell us the same information as a two-perimeter Rasch model. Second, our study implements structural equation modeling, which is in the same family as IFA/CFAs (Brown, 2006). Accordingly, we wish to keep our analyses within the same theoretical framework. Third, IRT techniques such as Rasch measurement have strong requirements of the data structure, such as assuming all items have identical variance. Factor analysis does not have these same restrictions and is thus a more robust technique for our requirements (Beaujean, 2014). Additionally, our study investigates attitudes and beliefs, as opposed to test scores, making factor analysis an ideal technique (DeVellis, 2012).

Model fit was tested using multiple goodness of fit indicators. The Tucker-Lewis Index (TLI), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA) were examined. For both the TLI and CFI, a value ≥ .95 suggests a good model fit, and values ≥ .90 suggests an acceptable fit (Hu & Bentler, 1999). For RMSEA, values ≤ .06 suggest a good model fit and values ≤ .08 suggest acceptable fit (McDonald & Ho, 2002). Additionally, standardized factor loadings were examined to determine if any questions were not sufficiently loading onto the scale. The following scales were examined: performing science practices, scientist recognition, classroom climate, STEM identity, STEM motivation, and STEM career aspirations. Results of the measurement model demonstrated that the indicator variables loaded significantly onto their constructs for all scales (*p* < .001). However, standardized factor loadings and goodness of fit indicators varied by scale. More detail is given below for each scale.

Performing science practices factor loadings ranged from .75 to .91 (see Table S1), with indicators that suggested a good fit; χ2(35) = 256.23, *p* < .001, TLI = 1.0, CFI = 1.0, RMSEA = .08, 95% CI [0.07, 0.09]. A similarly good fit was observed for classroom climate, with standardized factor loadings ranging from .77 to .92 (see Table S1), and strong fit indicators; χ2(27) = 621.08, *p* < .001; TLI = .99, CFI = 1.0, RMSEA = .14, 95% CI[0.14, 0.15]. STEM career aspirations factor loadings ranged from .55 to .76, with good fit indicators; χ2(2) = 57.01, *p* < .001; TLI = .90, CFI = .97, RMSEA = .16, 95% CI[0.13, 0.20].

For the remaining scales, fit indicators were not as strong, resulting in the removal of one or two questions for each scale, determined by examining standardized factor loadings for each question. For the recognition as a scientist scale, two questions had factor loadings below .50 (“Assignments allowed me to demonstrate my ability to evaluate evidence to my teaching assistant [my classmates]”). Once removed, factor loadings ranged from .64 to .86 (see Table S1) and indicators suggested a good fit; χ2(14) = 176.24, *p* < .001; TLI = .99, CFI = 1.0, RMSEA = .11, 95% CI[0.09, 0.12]. Similarly, the STEM identity scale had two questions with standardized factor loadings which fell below .50 (“I think about being a STEM person” and “The things that I like to do in my spare time are similar to what most STEM students also like”). Once removed, factor loadings ranged from .55 to .61 (see Table S1), with indicators suggesting a good fit; χ2(9) = 112.18, *p* < .001; TLI = .95, CFI = .97, RMSEA = .11, 95% CI [0.09, 0.12]. The STEM motivation scale also had one question that fell below .50 (“How easily can you learn the material in your STEM classes?”). Once removed, factor loadings ranged from .55 to .72 (see Table S1), with good fit indicators; χ2(9) = 321.67, *p* < .001; TLI = .87, CFI = .92, RMSEA = .18, 95% CI [0.16, 0.20]. These adjusted scales were used for the remainder of analyses. Finally, due to concern over scale overlap between the confidence-related motivation scale items and the confidence-related classroom climate items, a factor analysis was performed containing these items. The analysis indicated two principal components, one with the motivation confidence items (loadings: .68-.78) and one with the classroom climate items (loadings: .89-.92). Given they were distinct, we elected to keep the items separated in their respective scales.

**Table S1**.

*Item Factor Analysis for All Scales (N = 1,079)*

|  |  |
| --- | --- |
| **Performing Science Practices Item Questions** | **Factor Loading1** |
| 1. Using evidence to propose a model or a hypothesis | .88 |
| 2. Getting practice constructing or revising hypotheses | .90 |
| 3. Proposing an experiment to test a hypothesis | .91 |
| 4. Evaluating an experimental design in relation to how it will test a hypothesis | .87 |
| 5. Evaluating the evidence that supports a scientific claim | .76 |
| 6. Using scientific concepts to explain an observation or findings from an experiment | .77 |
| 7. Gained a better understanding of how to read scientific research articles | .79 |
| 8. Developing and evaluating hypotheses with classmates | .91 |
| 9. Working with classmates to formulate an experimental design to test hypotheses | .91 |
| **Classroom Climate Scale Item Question** | **Factor Loading1** |
| 1. My class experiences have made STEM seem more interesting to me | .91 |
| 2. My class experiences have stimulated my enthusiasm for STEM | .91 |
| 3. My class experiences have made STEM seem more enjoyable | .92 |
| 4. My class experiences have made me feel more relaxed about learning STEM | .83 |
| 5. My class experiences have increased my confidence in my ability to do STEM | .89 |
| 6. My class experiences have made the idea of taking advanced STEM classes seem more possible | .88 |
| 7. My class experiences have made helped me see that many other students like STEM, just as I do | .84 |
| 8. My class experiences have given me an opportunity to meet STEM students who were more enjoyable to be with than many of the other students at college | .77 |
| 9. My class experiences have given me an opportunity to make friends with people who like STEM | .78 |
| 10. Discussing scientific ideas and possible explanations with classmates | .75 |
| **Recognition as a Scientist Scale Item Questions** | **Factor Loading2** |
| 1. Assignments allowed me to demonstrate my ability to evaluate evidence to my professor | .64 |
| 2. Assignments provided opportunities to demonstrate my reasoning skills to my professor | .86 |
| 3. My professor recognized my intellectual contributions to class discussions | .77 |
| 4. Assignments provided opportunities to demonstrate my reasoning skills to my teaching assistant | .85 |
| 5. My teaching assistant recognized my intellectual contributions to class discussions | .76 |
| 6. Assignments provided opportunities to demonstrate my reasoning skills to my classmates | .83 |
| 7. My classmates recognized my intellectual contributions to class discussions | .81 |
| **STEM Identity Item Questions** | **Factor Loading3** |
| 1. Being a STEM person is part of how I feel about myself | .60 |
| 2. Being a STEM person is part of my self-image | .60 |
| 3. I am like other STEM students | .60 |
| 4. I am representative of what it means to be a STEM student | .61 |
| 5. My personality and values are similar to most STEM students | .55 |
| 6. I feel like I belong with other STEM students | .58 |
| **STEM Motivation Scale Item Questions** | **Factor Loading4** |
| 1. In general, how confident are you in your ability to do well in STEM courses? | .61 |
| 2. How successful do you expect to be in your STEM classes? | .58 |
| 3. How confident are you about understanding the material in your STEM classes? | .55 |
| 4. How important to you are your STEM classes? | .65 |
| 5. How much do you value your STEM classes? | .72 |
| 6. How useful do you consider your STEM classes? | .57 |

*Note.* All factor loadings significantly loaded onto constructs (*p* < .001).

1No items removed due to factor loadings < .50.

2 Two questions were removed due to low factor loadings (“Assignments allowed me to demonstrate my ability to evaluate evidence to my teaching assistant” and “Assignments allowed me to demonstrate my ability to evaluate evidence to my classmates”).

3 Two questions were removed due to low factor loadings (“I think about being a STEM person” and “The things that I like to do in my spare time are similar to what most STEM students also like”).

4 One question was removed due to low factor loadings (“How easily can you learn the material in your STEM classes?”).