# Supplemental Material

*CBE—Life Sciences Education*

Estrada *et al.*

Supplemental Materials

# A Longitudinal Study of How Quality Mentorship and Research Experience Integrate Underrepresented Minorities into STEM Careers

Mica Estrada

University of California, San Francisco

Department of Social &Behavioral Sciences,

Institute for Health & Aging

3333 California Street, Suite 340, San Francisco, CA 94118.

mica.estrada@ucsf.edu.

Paul R. Hernandez

West Virginia University

P. Wesley Schultz

California State University, San Marcos

***Mentoring support scale***

As noted in the main text, we adapted administered a 9-item version of the mentoring support scale developed by Dreher and Ash (1990). Participants read the following instructions and completed the following 9-items.

Instructions: Please answer the following questions about your faculty mentor. To what extent does your mentor provide each of the following? Please select the best answer on the scale from 1 (not at all) to 5 (to a very large extent).

1. To what extent has your mentor discussed your questions or concerns regarding feelings of competence, commitment to advancement or relationships with peers?
2. To what extent has your mentor conveyed empathy for concerns or feelings you have discussed with him or her?
3. To what extent has your mentor encouraged you to talk openly about anxieties and fears?
4. To what extent has you mentor shared personal experiences with you?
5. To what extent has your mentor helped you finish assignments/tasks or meet deadlines that otherwise would have been difficult to complete?
6. To what extent has your mentor helped you improve your writing skills?
7. To what extent has your mentor given you challenging assignments that present opportunities to learn new skills?
8. To what extent has your mentor helped you meet people elsewhere?
9. To what extent has your mentor helped you meet other people in your field at the university?

Note that items 1-4 reflect psychosocial support, items 5-7 reflect instrumental support, and items 8-9 reflect networking support.

***Intra-class Correlations (ICCs)***

ICCs were calculated for all of the TIMSI variables as well as the outcome variables. As noted in the main text, the ICCs were exceedingly small, resulting in multi-level SEMs that were unstable. The ICCs were as follows:

- Science Efficacy Junior-Fall = .01
- Science Efficacy Junior-Spring = .002
- Science Efficacy Senior-Fall = .01
- Science Efficacy Senior-Spring = .01
- Science Identity Junior-Fall = .05
- Science Identity Junior-Spring= .04
- Science Identity Senior-Fall = .02
- Science Community Values Senior- Fall = .02
- STEM career = .04
- Medical career = .04
- Other career = .01

***Preliminary growth model analytic description***

Prior to addressing our research questions and hypotheses, we estimated a baseline structural equation model to assess and describe the levels and development of science efficacy, science identity, and science community values (i.e., measures of integration into the scientific community). As shown in Figure S1, latent growth curve models were fit to the science efficacy and science identity data, in addition to estimating the level of science community values in senior year. Science efficacy was measured fall and spring semesters of both junior and senior years. However, science identity was only measured fall and spring of junior year and fall of senior year. In addition, science community values were only measured in the fall semester of senior year. These differences in measurement were due to the fact that some scales were rotated in or out of the survey administration protocol. The growth curve models were specified so that the average level of science efficacy and identity in fall semester of senior year across all students were estimated by the intercept of the growth models. The level of science community values was estimated as the mean across student in the fall semester of senior year. In addition, the model was specified to capture the average per-semester linear change of science efficacy and identity over time across all students by estimating the linear growth slope.

The analysis revealed that the model provided good fit to the data, see Figure S1. An inspection of the parameter estimates revealed that the level (i.e., intercept) of science efficacy in the fall semester of senior year were moderately high (i.e., 4.02 on a 1-5 scale). In addition, the growth model indicated that, on average, student science efficacy grew by a small but statistically significant amount each passing semester. However, there was a statistically significant amount of variability around the average level and average growth rate of science efficacy.

Similarly, the model indicated that the average level of science identity in the fall semester of senior year (i.e., intercept) was moderately high with significant variability around the intercept. However, unlike science efficacy, the science identity linear growth slope was essentially zero indicating no average change per semester. In addition, the model showed that the growth rate did not vary significantly across individuals. Finally, the model indicated that the average level of science community values in the fall semester of senior year was moderately high with significant variability around the mean.

Finally, an inspection of the correlations revealed that the levels (i.e., intercepts) of science efficacy, identity, and community values in fall semester of senior year were moderately to strongly and positively correlated, see Figure S1.

***Analytic approach to nested structure***

We recognize that these data have a nested structure of occasions within students within institutions. We assessed variability at the institutional level by calculating the intra-class correlation coefficients (ICCs) for each variable in our model using a 3-level multilevel model. The ICCs were exceedingly small at the institution level (ICC $M$ = .026, $SD$ = .019), indicating very little campus level variability (see Supplemental materials for complete details). In fact, attempts to model the data in multilevel SEM (i.e., 3-level model) consistently failed to converge. Therefore, we proceeded with a typical SEM analysis (i.e., 2-level model).

**SUPPLEMENTAL TABLE 1. Correlation matrix and descriptive statistics for the variables used in all models**[a]

| | Variable | N | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | STEM career | 694 | 0.44 | 0.50 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | Medical/Clinical career | 694 | 0.28 | 0.45 | -.94 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | Other career | 694 | 0.28 | 0.45 | -.94 | -.88* | -- | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Science Efficacy (Jr.-fall) | 216 | 4.00 | 0.65 | .13* | .02 | -.18* | *.90* | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Science Efficacy (Jr.-spring) | 511 | 3.98 | 0.67 | .00 | .14* | -.14* | .66* | *.90* | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Science Efficacy (Sr.-fall) | 447 | 3.98 | 0.66 | .18* | -.12* | -.08 | .75* | .61* | *.89* | | | | | | | | | | | | | | | | | | | | | | |
| 7 | Science Efficacy (Sr.-spring) | 747 | 4.07 | 0.62 | .13* | -.06 | -.09 | .43* | .54* | .68* | *.89* | | | | | | | | | | | | | | | | | | | | | |
| 8 | Science Identity (Jr.-fall) | 212 | 3.78 | 0.84 | .27* | -.10 | -.20* | .45* | .38* | .35* | .39* | *.83* | | | | | | | | | | | | | | | | | | | | |
| 9 | Science Identity (Jr.-spring) | 397 | 3.77 | 0.79 | .27* | -.03 | -.30* | .33* | .45* | .49* | .36* | .58* | *.81* | | | | | | | | | | | | | | | | | | | |
| 10 | Science Identity (Sr.-fall) | 436 | 3.73 | 0.91 | .28* | -.12* | -.20* | .30* | .37* | .57* | .52* | .49* | .74* | *.87* | | | | | | | | | | | | | | | | | | |
| 11 | Science Community Values (Sr.-fall) | 324 | 4.78 | 0.99 | .31* | -.14* | -.20* | .40* | .41* | .60* | .48* | .46* | .43* | .66* | *.89* | | | | | | | | | | | | | | | | | |
| 12 | Mentor Quality (Jr. to Sr.) | 602 | 3.90 | 0.77 | .09 | -.03 | -.07 | .07 | .09 | .05 | .21* | .29* | .18* | .09 | .21* | *.78* | | | | | | | | | | | | | | | | |
| 13 | Carnegie=Research Univ. (Very High Research Activity) | 1015 | 0.04 | 0.20 | -.11* | .14* | -.05 | -.02 | -.01 | -.01 | -.04 | -.08 | -.12* | -.12* | -.01 | -.12* | -- | | | | | | | | | | | | | | | |
| 14 | Carnegie=Research Univ. (High Research Activity) | 1015 | 0.24 | 0.42 | .07 | -.04 | -.05 | -.06 | .05 | -.02 | .07 | .08 | .16* | .00 | -.04 | .00 | -.11* | -- | | | | | | | | | | | | | | |
| 15 | Carnegie=Doctoral/Research Univ. | 1015 | 0.06 | 0.22 | .00 | -.04 | .04 | .00 | -.01 | .05 | .01 | .01 | .02 | .05 | .03 | -.07 | -.05 | -.14* | -- | | | | | | | | | | | | | |
| 16 | Carnegie=Master's (Larger Programs) | 1015 | 0.31 | 0.46 | -.02 | .03 | -.01 | .02 | -.03 | -.04 | -.07 | .06 | -.03 | .05 | -.10 | -.04 | -.13* | -.37* | -.16* | -- | | | | | | | | | | | | |
| 17 | Carnegie=Master's (Medium Programs) | 1015 | 0.13 | 0.35 | .09 | -.11* | -.01 | -.02 | .02 | .12* | .02 | -.09 | -.01 | .04 | .12 | .05 | -.08 | -.22* | -.10 | -.26* | -- | | | | | | | | | | | |
| 18 | Carnegie=Master's (Smaller Programs) | 1015 | 0.06 | 0.22 | .03 | -.06 | .02 | .03 | -.03 | .01 | .05 | .00 | .01 | .02 | .03 | .00 | -.05 | -.14* | -.06 | -.16* | -.10 | -- | | | | | | | | | | |
| 19 | Carnegie=Baccalaureate College (Arts & Sciences) | 1015 | 0.12 | 0.33 | -.09 | .06 | .05 | .14* | -.02 | -.06 | -.02 | -.02 | -.14* | -.08 | .07 | .09 | -.08 | -.21* | -.09 | -.25* | -.15* | -.09 | -- | | | | | | | | | |
| 20 | Ethnic=Asian | 1015 | 0.05 | 0.20 | .05 | .00 | -.07 | -.13* | -.06 | -.06 | -.07 | -.01 | -.01 | -.08 | -.16* | -.04 | .03 | -.06 | -.03 | .11* | .00 | -.05 | .01 | -- | | | | | | | | |
| 21 | Ethnic=Hawaiian/Pacific Islander | 1015 | 0.02 | 0.14 | .00 | .03 | -.03 | .10 | .01 | -.04 | -.05 | .01 | -.03 | -.05 | .06 | .02 | .01 | -.07 | -.03 | .01 | -.01 | -.03 | .14* | -.03 | -- | | | | | | | |
| 22 | Ethnic=Hispanic/Latino/Latina | 1015 | 0.40 | 0.49 | .07 | .03 | -.11* | -.07 | .02 | .06 | -.01 | -.05 | .10 | .13* | .11* | -.04 | .10 | .34* | -.14* | .04 | .05 | -.17* | -.27* | -.18* | -.11* | -- | | | | | | |
| 23 | Ethnic=Native American/Alaskan Native | 1015 | 0.01 | 0.10 | .02 | -.05 | .02 | .01 | -.03 | .00 | .05 | -.02 | .02 | -.02 | .06 | -.03 | -.02 | -.01 | .01 | .01 | .00 | -.03 | .03 | -.03 | -.02 | -.10 | -- | | | | | |
| 24 | Ethnic=White/non-Hispanic | 1015 | 0.05 | 0.22 | -.02 | .00 | .02 | .07 | .01 | .04 | .05 | -.03 | .08 | .01 | -.02 | -.03 | .02 | -.05 | -.02 | .09 | .04 | -.04 | -.05 | -.05 | -.03 | -.19* | -.03 | -- | | | | |
| 25 | Gender = Male | 1015 | 0.28 | 0.45 | .08 | .00 | -.10 | .09 | .03 | .15* | .07 | .11* | .07 | .01 | .02 | -.04 | -.01 | .00 | -.02 | .04 | .01 | -.05 | -.04 | .05 | .07 | .04 | -.02 | .13* | -- | | | |
| 26 | Baseline GPA | 1015 | 3.32 | 0.41 | .01 | .18* | -.19* | -.09 | .02 | .01 | .02 | -.07 | .05 | .03 | .06 | .03 | .08 | .11* | .00 | -.07 | .01 | -.02 | -.11* | .06 | -.09 | .09 | .00 | .10 | .01 | -- | | |
| 27 | Research Experience (1-semenster) | 651 | 0.31 | 0.46 | .01 | -.02 | .00 | -.02 | -.01 | .06 | .04 | -.08 | -.08 | .02 | .03 | .01 | -.01 | -.01 | .01 | .01 | .03 | -.06 | .02 | -.09* | -.03 | .04 | .01 | .08* | .05 | .01 | -- | |
| 28 | Research Experience (2-sememsters) | 651 | 0.25 | 0.43 | .26** | -.13* | -.16* | .10 | .15** | .16* | .15* | .19* | .23* | .29* | .30* | -.05 | .00 | .19* | -.11* | .10* | -.04 | -.01 | .03 | .05 | .07 | .02 | -.01 | -.01 | .12* | .26* | -.39* | -- |

a. Values on diagonal are reliability (Crombach's alpha) estimates; values on off diagonals are correlation coefficients; Mean values for dichotomous variables (i.e., career outcomes, Carnegie classifications, ethnicity, & gender) represent proportion of the sample in the focal category
Significance codes: *$p \leq .05$

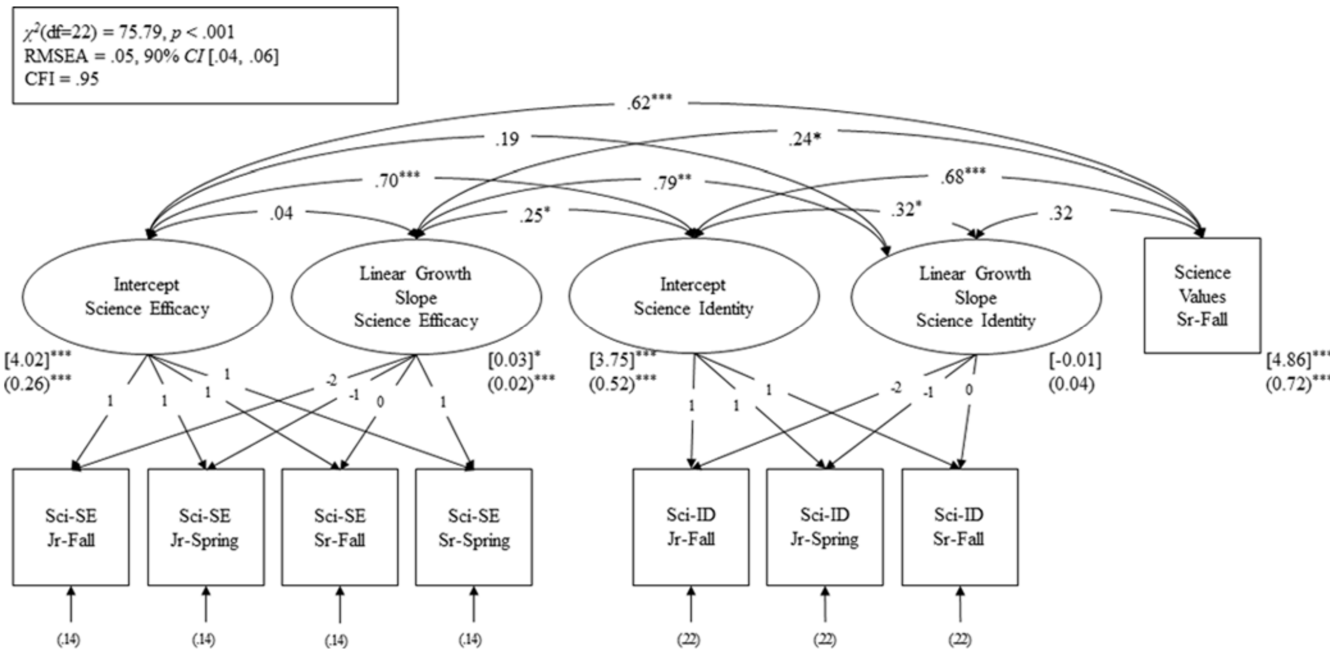


**Figure S1.** TIMSI variables linear growth curve models

Figure notes: Values inside [brackets] represent intercept or mean; values inside (parentheses) represent variances (or residual variance); values on curved double-headed lines represent correlation coefficients. Boxes represent observed variables and ellipses represent latent variables; Sci-SE = Science Self-Efficacy; Sci-ID = Science Identity; Jr = Junior year; Sr = Senior Year; Fall = fall semester; Spring = spring semester; The paths/arrows leading from each latent slope to the observed indicators (i.e., boxes) are centered on fall semester of senior year and change by one unit before/after fall semester of senior year. This coding scheme allows the linear growth slope to quantify the average linear change per semester of time and ensures that the common intercept/level for each construct is fall semester of senior year.

*p<.05, **p<01, ***p<.001