Supplemental Material

CBE—Life Sciences Education

Tuma and Dolan

Supplemental Online Materials for:

What makes a good match?
Predictors of quality mentorship among doctoral students

Trevor T. Tuma* & Erin L. Dolan

* Corresponding author. Email: trevor.tuma@uga.edu

Table of Contents

Scales used to measure latent variables

Culturally aware mentoring behaviors
Byars-Winston, A., & Butz, A. R. (2021). Measuring research mentors' cultural diversity awareness for race/ethnicity in STEM: Validity evidence for a new scale. *CBE—Life Sciences Education*, *20*(2), ar15.

Instructions: Please indicate the extent to which you agree or disagree with each of these statements.
Response Scale: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree, 6 = Prefer not to respond

Culturally aware mentoring behaviors	1. My mentor created opportunities for me to bring up issues of race/ethnicity as they arose.
	2. My mentor encouraged me to think about how the research related to my own lived experience.
	3. My mentor was willing to discuss race and ethnicity, even if it may have been uncomfortable for them.
	4. My mentor raised the topic of race/ethnicity in our research mentoring relationship when it was relevant.
	5. My mentor approached the topic of race/ethnicity with me in a respectful manner.

Deep-level similarity
Ensher, E. A., Grant-Vallone, E. J., & Marelich, W. D. (2002). Effects of perceived attitudinal and demographic similarity on protégés' support and satisfaction gained from their mentoring relationships. *Journal of Applied Social Psychology*, *32*(7), 1407-1430.

Instructions: Please indicate the extent to which you agree or disagree with each of these statements.
Response Scale: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree, 6 = Prefer not to respond

Deep-level similarity	1. My mentor and I see things in the same way.
	2. My mentor is similar to me in terms of our outlook and perspectives.
	3. My mentor and I are alike in a number of areas.
	4. My mentor and I analyze problems in a similar way.
	5. My mentor and I have similar values about work.
	6. My mentor and I have similar values about life in general.
	7. My mentor and I are more similar than dissimilar in important ways.

Career support[1]
Tuma TT, Adams, JD, Choi S, & Dolan EL (unpublished data) Measuring Negative Mentoring Experiences: Development & Nomological Validation of the Mentoring Experiences in Research & Graduate Education (MERGE) Scale[1]

Instructions: Please indicate the extent to which you agree or disagree with each of these statements.
Response Scale: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree, 6 = Prefer not to respond

[1] This measure is currently being developed to measure doctoral students' negative mentoring experiences. Validity and reliability evidence based on the test content, response process, internal structure, and relationships with other variables has been collected and the resulting measurement validation manuscript is in preparation.

Career support	1. My mentor is willing to give me feedback on my research.
	2. My mentor advocates on my behalf.
	3. My mentor protects me from others who might cause me professional harm.
	4. My mentor has little interest in my career advancement. (R)
	5. My mentor is reluctant to let me present my research at conferences. (R)
	6. My mentor offers useful advice for achieving my career goals.
	7. My mentor prioritizes publishing my research.
	8. My mentor helps me identify ways to network.
	9. My mentor makes sure I have sufficient funding to do my research.
	10. My mentor helps me prepare for important milestones in my degree.

Psychosocial support[1]
Tuma TT, Adams, JD, Choi S, & Dolan EL (unpublished data) Measuring Negative Mentoring Experiences: Development & Nomological Validation of the Mentoring Experiences in Research & Graduate Education (MERGE) Scale.

Instructions: Please indicate the extent to which you agree or disagree with each of these statements.
Response Scale: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree, 6 = Prefer not to respond

Psychosocial support	1. My mentor encourages me.
	2. My mentor values me as a person.
	3. My mentor checks in about my well-being.
	4. My mentor is a role model for me.
	5. My mentor makes me feel accepted.
	6. My mentor empathizes when I am struggling.
	7. My mentor is understanding when I experience difficulties.
	8. My mentor tells me when they think I have done a good job.

Relationship quality[1]
Tuma TT, Adams, JD, Choi S, & Dolan EL (unpublished data) Measuring Negative Mentoring Experiences: Development & Nomological Validation of the Mentoring Experiences in Research & Graduate Education (MERGE) Scale.

Instructions: Please indicate the extent to which you agree or disagree with each of these statements.
Response Scale: 1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree, 6 = Prefer not to respond

Relationship quality	1. My mentor and I do not like each other. (R)
	2. My mentor and I have a difficult relationship. (R)
	3. My mentor and I can talk about things other than work tasks.
	4. My mentor and I have a tense relationship. (R)

Results from measurement model fitting and model modifications

We began the first phase of our SEM by evaluating the relationships between our measured indicators and the underlying latent variables (i.e., measurement model). We performed confirmatory factor analysis (CFA) using the R software (R Core Team, 2016) for statistical computing and the 'lavaan' package (Rosseel, 2012) for cultural aware mentoring, deep-level similarity, career support, psychosocial support, and relationship quality. We used robust maximum likelihood (MLR) estimation to correct for potential non-normality in our data.

Then, we assessed how well our measurement models reproduced their variance-covariance matrices using several goodness of fit indices. First, we examined the model's chi-square test ($\chi 2$) of goodness of fit and its degrees of freedom and *p*-value to evaluate the discrepancy between a hypothesized model and the data. It is generally agreed upon best practice to report the chi-square test despite its limitations (e.g., sensitivity to sample size, unrealistic null hypothesis in a population). Consistent with current measurement standards, we also evaluated multiple model fit indices (e.g., absolute, parsimonious, incremental) relative to recommended "cut-off" values (Hu & Bentler, 1999). The comparative fit index (CFI) and Tucker-Lewis index (TLI) are both incremental fit measures and values $\geq$.90 indicate acceptable data-model fit. In addition, we also examined the root mean square error of residuals (RMSEA) value. RMSEA is a parsimony-adjusted fit index and approximates how well the model estimates the population covariance matrix while penalizing more complex models, with higher RMSEA values indicating poorer data-model fit. We used 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre model fit. Finally, we evaluated the standardized root mean square residual (SRMR) which examines the standardized difference between the observed correlation and the predicted correlation, thus giving further insight into mis-specified covariance structures. The SRMR is an absolute fit indice, with higher values (< 0.08) indicating poor model fit. For all model fit indices, values close to the cut-off are categorized as indicating fair fit. In sum, high values ($\geq$.90) of the CFI and TLI and low values of the RMSEA and SRMR provide evidence that the hypothesized measurement models are plausible explanations of the data.

Finally, in addition to traditional null hypothesis testing, we used equivalence testing (Marcoulides & Yuan 2017; Peugh & Feldon, 2020; Yuan et al., 2016). We calculated adjusted, or "T-size" fit statistics, to compare the amount of misspecification in our model to a tolerable size of specification with adjusted cutoffs. Adjusted cutoffs and T-size fit statistics are reported below for CFI_T and $RMSEA_T$.

We report coefficient Omega values, which is considered to be a more sensible indicator of internal consistency than Cronbach's alpha, as an indicator of reliability (Dunn et al., 2014). We also encourage interested readers to consult the items comprising the scales to judge their face equivalence (i.e., the extent to which the items appear to elicit the same underlying latent variable).

Cultural aware mentoring behaviors

We ran a CFA with cultural diversity awareness behaviors indicated by five items. The majority of the model fit statistics indicated excellent fit, $\chi 2$ (5) = 21.552 (p<0.001), CFI = 0.982, TLI = 0.965, SRMR = 0.022, with the exception of RMSEA = 0.099 90% CI [0.059, 0.144]) which indicated inadequate fit. Furthermore, the CFI_T demonstrated close fit and the $RMSEA_T$ indicated fair fit with adjusted fit values [A], $\chi 2$ (5) = 21.552 (p<0.001), CFI_T =0.951, $RMSEA_T$ =0.111. Therefore, we opted to proceed without adjusting the measurement model. The scale demonstrated high internal consistency (α=0.86) and (ω=0.90).

[A] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.832, "mediocre" = 0.832 – 0.860, "fair" = 0.860 – 0.905, "close" = 0.905 – 0.968 and "excellent" are $\geq$ 0.968

and for $RMSEA_T$: "poor" $\geq$ 0.137, "mediocre" = 0.117 – 0.137, "fair" = 0.088 – 0.117, "close" = 0.059 – 0.088, and "excellent" are $\leq$ 0.059.

Deep-level similarity

We ran a CFA with deep-level similarity indicated by seven items. The majority of the model fit statistics indicated excellent fit, $\chi 2$ (14) = 139.466 (p<0.001), CFI = 0.945, TLI = 0.917, SRMR = 0.034, with the exception of RMSEA = 0.136 90% CI [0.116, 0.157]) which indicated inadequate fit. Furthermore, the CFI_T demonstrated fair fit and the $RMSEA_T$ indicated poor fit with adjusted fit values [B], $\chi 2$ (14) = 139.466 (p<0.001), CFI_T =0.903, $RMSEA_T$ =0.145. To determine the source of misfit, we checked for correlated residuals using modification indices as a guide. Using a backwards selection process, we first identified the largest modification index (MI) value and ensured that the potential parameter modification was theoretically supported. We then respecified the model with the additional correlated residuals and examined the goodness of fit indices of the revised model. Our modification indices recommended allowing the residual values of item 6 and item 7 to correlate (MI=45.192), item 3 and item 7 (MI=32.005), and item 5 and item 6 (MI=23.501). Adding these paths to the model improved model fit $\chi 2$ (11) = 57.850 (p< 0.001), CFI = 0.980, TLI = 0.962, SRMR = 0.023, RMSEA = 0.092, 90% CI [0.069 - 0.115]). These modifications were further supported with CFI_T demonstrating close fit and the $RMSEA_T$ indicating mediocre fit with adjusted fit values [C], $\chi 2$ (11) = 57.850 (p< 0.001)), CFI_T =0.955 $RMSEA_T$ =0.109. The scale demonstrated high internal consistency (α=0.92) and (ω=0.94).

[B] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.848, "mediocre" = 0.848 – 0.875, "fair" = 0.875 – 0.915, "close" = 0.915 – 0.972 and "excellent" are $\geq$ 0.972 and for $RMSEA_T$: "poor" $\geq$ 0.121, "mediocre" = 0.101 – 0.121, "fair" = 0.072 – 0.101, "close" = 0.042 – 0.072, and "excellent" are $\leq$ 0.042.

[C] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.848, "mediocre" = 0.848 – 0.875, "fair" = 0.875 – 0.915, "close" = 0.915 – 0.972 and "excellent" are $\geq$ 0.972 and for $RMSEA_T$: "poor" $\geq$ 0.124, "mediocre" = 0.104 – 0.124, "fair" = 0.075 – 0.141, "close" = 0.046 – 0.075, and "excellent" are $\leq$ 0.046.

Career support

We first ran a CFA with perceived career support indicated by ten items. The fit indices of the model were acceptable, $\chi 2$ (35) = 186.86 (p<0.001), CFI = 0.927, TLI = 0.906, SRMR = 0.041, with the exception of RMSEA = 0.099, 90% CI [0.085, 0.113]) which indicated inadequate fit. Furthermore, CFI_T demonstrated mediocre fit and $RMSEA_T$ indicated mediocre fit with adjusted fit values [D], $\chi 2$ (20) = 128.89 (p<0.001), CFI_T =0.881, $RMSEA_T$ =0.100. To determine the source of misfit, we checked for correlated residuals using modification indices as a guide. Using a backwards selection process, we first identified the largest modification index (MI) value and ensured that the potential parameter modification was theoretically supported. We then respecified the model with the additional correlated residuals and examined the goodness of fit indices of the revised model. Our modification indices recommended allowing the residual values of item 2 and item 3 to correlate (MI=88.176), then item 1 and item 8 (MI=31.451). Adding these paths to the model improved model fit $\chi 2$ (33) = 92.240 (p<0.001), CFI = 0.971, TLI = 0.961, SRMR = 0.032, RMSEA = 0.064, 90% CI [0.048 0.079]). These modifications were further supported with CFI_T demonstrating close fit and the $RMSEA_T$ indicating fair fit with adjusted fit values [E], $\chi 2$ (33) = 92.240 (p<0.001), CFI_T =0.943, $RMSEA_T$ =0.070. The scale demonstrated high internal consistency (α=0.89) and (ω=0.92).

[D] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.860, "mediocre" = 0.860 – 0.885, "fair" = 0.885 – 0.923, "close" = 0.923 – 0.976, and "excellent" are $\geq$ 0.976 and $RMSEA_T$: "poor" $\geq$ 0.112, "mediocre" = 0.092 – 0.112, "fair" = 0.064 – 0.092, "close" = 0.032 – 0.064, and "excellent" are $\leq$ 0.032.

[E] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.861, "mediocre" = 0.861 – 0.885, "fair" = 0.885 – 0.923, "close" = 0.923 – 0.976, and "excellent" are $\geq$ 0.976 and for $RMSEA_T$: "poor" $\geq$ 0.112, "mediocre" = 0.093 – 0112, "fair" = 0.064 – 0.093, "close" = 0.033 – 0.064, and "excellent" are $\leq$ 0.033.

Psychosocial support

We ran a CFA with perceived psychosocial support indicated by eight items. The majority of the model fit statistics indicated excellent fit, $\chi2$ (20) = 149.094 (p<0.001), CFI = 0.948, TLI = 0.928, SRMR = 0.028, with the exception of RMSEA = 0.137, 90% CI [0.116, 0.157]) which indicated inadequate fit. Furthermore, the CFI_T demonstrated fair fit and the $RMSEA_T$ indicated poor fit with adjusted fit values [F], $\chi2$ (20) = 149.10 (p<0.001), CFI_T =0.916, $RMSEA_T$ =0.123. To determine the source of misfit, we checked for correlated residuals using modification indices as a guide. Using a backwards selection process, we first identified the largest modification index (MI) value and ensured that the potential parameter modification was theoretically supported. We then respecified the model with the additional correlated residuals and examined the goodness of fit indices of the revised model. Our modification indices recommended allowing the residual values of item 6 and item 7 to correlate (MI=105.537), then item 1 and item 2 (MI=27.226), and then item 1 and item 8 (27.981). Adding these paths to the model improved model fit $\chi2$ (17) = 57.264 (p<0.001), CFI = 0.985, TLI = 0.975, SRMR = 0.019, RMSEA = 0.080, 90% CI [0.058 - 0.103]). These modifications were further supported with CFI_T demonstrating close fit and the $RMSEA_T$ indicating fair fit with adjusted fit values [G], $\chi2$ (33) = 92.240 (p<0.001), CFI_T =0.966, $RMSEA_T$ =0.083. The scale demonstrated high internal consistency (α=0.95) and (ω=0.96).

[F] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.853, "mediocre" = 0.853 – 0.879, "fair" = 0.879 – 0.918, "close" = 0.918 – 0.974, and "excellent" are $\geq$ 0.974 and for $RMSEA_T$: "poor" $\geq$ 0.117, "mediocre" = 0.097 – 0.117, "fair" = 0.069 – 0.097, "close" = 0.038 – 0.069, and "excellent" are $\leq$ 0.038.

[G] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.854, "mediocre" = 0.854 – 0.879, "fair" = 0.879 – 0.918, "close" = 0.918 – 0.974, and "excellent" are $\geq$ 0.974 and for $RMSEA_T$: "poor" $\geq$ 0.118, "mediocre" = 0.099 – 0.118, "fair" = 0.070 – 0.099, "close" = 0.040 – 0.070, and "excellent" are $\leq$ 0.040.

Relationship quality

We ran a CFA with perceived relationship quality indicated by four items. The majority of the model fit statistics indicated excellent fit, $\chi2$ (2) = 19.856 (p<0.001), CFI = 0.989, TLI = 0.967, SRMR = 0.018, with the exception of RMSEA = 0.130, 90% CI [0.082, 0.185]) which indicated inadequate fit. Furthermore, the CFI_T demonstrated close fit and the $RMSEA_T$ indicated poor fit with adjusted fit values [H], $\chi2$ (2) = 19.856 (p<0.001), CFI_T =0.946, $RMSEA_T$ =0.178. To determine the source of misfit, we checked for correlated residuals using modification indices as a guide. Using a backwards selection process, we first identified the largest modification index (MI) value and ensured that the potential parameter modification was theoretically supported. We then respecified the model with the additional correlated residuals and examined the goodness of fit indices of the revised model. Our modification indices recommended allowing the residual values of item 2 and item 4 to correlate (MI=21.326). Adding this path to the model improved model fit $\chi2$ (1) = 0.365 (p = 0.546), CFI = 1.000, TLI = 1.002, SRMR = 0.002, RMSEA = 0.000, 90% CI [0.000 - 0.098]). These modifications were further supported with CFI_T demonstrating excellent fit and the $RMSEA_T$ indicating excellent fit with adjusted fit values [I], $\chi2$ (1) = 0.365, (p<0.001), CFI_T =0.991 $RMSEA_T$ =0.09. The scale demonstrated high internal consistency (α=0.89) and (ω=0.91).

[H] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.815, "mediocre" = 0.815 – 0.846, "fair" = 0.846 – 0.894, "close" = 0.894 – 0.963, and "excellent" are $\geq$ 0.963 and for $RMSEA_T$: "poor" $\geq$ 0.159, "mediocre" = 0.139 – 0.159, "fair" = 0.110 – 0.139, "close" = 0.082 – 0.110, and "excellent" are $\leq$ 0.082.

[I] The rescaled fit statistic values for the measurement model indicated for CFI_T: "poor" $\leq$ 0.807, "mediocre" = 0.807 – 0.838, "fair" = 0.838 – 0.888, "close" = 0.888 – 0.960, and "excellent" are $\geq$ 0.960 and for $RMSEA_T$: "poor" $\geq$ 0.183, "mediocre" = 0.163 – 0.183, "fair" = 0.133 – 0.163, "close" = 0.133– 0.108, and "excellent" are $\leq$ 0.108.

TABLE S1. Doctoral student-faculty member characteristics in the analyses by gender and race and ethnicity match.

Dyad composition	Total
Gender dyads	558
Similar gender dyad	280 (50%)
Dissimilar gender dyad	278 (50%)
Race and ethnicity dyads	517
Similar race and ethnicity dyad	246 (48%)
Dissimilar race and ethnicity dyad	271 (52%)

TABLE S2. Characteristics of gender composition mentee-mentor dyads (total *n* = 558)

Dyad composition	Total
Woman mentee-woman mentor	126 (23%)
Man mentee-man mentor	154 (28%)
Woman mentee-man mentor	199 (36%)
Man mentee-woman mentor	66 (12%)
Non binary mentee-woman mentor	5 (<1%)
Non binary mentee-man mentor	8 (1%)

TABLE S3. Descriptive statistics of indicators of quality mentorship by gender dyad composition

Mentee	*n*	Mentor	*n*	Career Support		Psychosocial Support		Relationship Quality	
				M	***SD***	***M***	***SD***	***M***	***SD***
Woman	325	Woman	126	3.90	0.68	3.79	1.03	4.05	0.97
		Man	199	3.79	0.84	3.80	1.04	4.07	0.93
Man	220	Woman	66	3.72	0.92	3.61	1.19	3.83	1.08
		Man	154	3.88	0.75	3.75	0.97	4.02	0.92
Non-binary	13	Woman	5	3.48	1.01	3.52	1.43	3.95	1.11
		Man	8	3.30	0.80	3.20	1.32	3.50	1.08

TABLE S4. Characteristics of racial and ethnic composition mentee-mentor dyads reported in Figure 3 of main text (total *n* = 617)

Total Dyads	Dyad composition	Total	Dyad composition	Total	Dyad composition	Total
127	Asian mentee & Asian mentor	27 (21%)	Asian mentee & racially minoritized mentor	10 (8%)	Asian mentee & White mentor	90 (71%)
84	Hispanic mentee & Hispanic mentor	9 (11%)	Hispanic mentee & Non-Hispanic mentor	75 (89%)		
77	Racially minoritized mentee & racially minoritized mentor	6 (8%)	Racially minoritized mentee & Asian mentor	11 (14%)	Racially minoritized mentee & White mentor	60 (78%)
329	White mentee & White mentor	281 (85%)	White mentee & Asian mentor	37 (11%)	White mentee & racially minoritized mentor	11 (4%)

TABLE S5: Descriptive statistics of mentoring support and relationship quality by racial/ethnic composition of dyads.

Mentee	*n*	Mentor	*n*	Career Support		Psychosocial Support		Relationship Quality	
				M	***SD***	***M***	***SD***	***M***	***SD***
American Indian or Alaskan Native	10	American Indian or Alaskan Native	0						
		Asian	2	4.95	0.07	4.94	0.09	4.88	0.17
		Black or African American	0						
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	0						
		North African or Middle Eastern	0						
		White	7	3.69	0.93	3.80	1.23	3.78	1.19
Asian	131	American Indian or Alaskan Native	3	4.50	0.36	4.54	0.38	4.50	0.50
		Asian	27	3.69	0.90	3.63	1.16	3.85	1.07
		Black or African American	2	4.85	0.07	5.00	0.00	4.87	0.18
		Hawaiian or Native Pacific Islander	1	4.80	0.00	5.00	0.00	5.00	0.00
		Hispanic or Latinx/Latine	4	4.10	0.47	4.10	0.84	4.25	0.54
		North African or Middle Eastern	4	4.30	0.44	4.28	0.59	4.38	0.78
		White	90	3.77	0.70	3.96	0.86	4.10	0.78
Black or African American	52	American Indian or Alaskan Native	0						
		Asian	7	4.16	0.74	4.18	0.96	4.50	0.46
		Black or African American	3	3.90	0.95	4.00	0.88	4.08	0.88
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	4	4.05	0.74	4.12	1.07	4.56	0.32
		North African or Middle Eastern	0						
		White	39	3.88	0.77	3.80	0.99	4.00	1.02
Hawaiian or Native Pacific Islander	3	American Indian or Alaskan Native	0						
		Asian	0						
		Black or African American	0						
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	0						
		North African or Middle Eastern	0						
		White	3	3.46	1.10	3.58	1.06	3.66	1.28

TABLE S5 cont.

Mentee	*n*	Mentor	*n*	Career Support		Psychosocial Support		Relationship Quality	
				M	***SD***	***M***	***SD***	***M***	***SD***
Hispanic or Latinx/Latine	86	American Indian or Alaskan Native	0						
		Asian	8	4.07	0.62	3.89	0.75	4.19	0.95
		Black or African American	3	4.53	0.32	3.96	0.63	3.66	1.70
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	9	3.72	0.76	3.51	0.95	3.80	0.81
		North African or Middle Eastern	1	4.70	0.00	4.37	0.00	4.75	0.00
		White	66	3.68	0.86	3.50	1.21	3.89	1.02
North African or Middle Eastern	15	American Indian or Alaskan Native	0						
		Asian	2	3.85	0.92	3.50	1.06	4.00	1.41
		Black or African American	1	1.90	0.00	1.00	0.00	1.5	0.00
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	1	1.90	0.00	1.00	0.00	1.5	0.00
		North African or Middle Eastern	2	3.5	0.42	2.62	0.17	3.62	0.88
		White	11	3.59	0.84	3.19	1.19	3.63	1.12
White	339	American Indian or Alaskan Native	0						
		Asian	37	3.76	0.77	3.58	1.10	3.98	0.98
		Black or African American	6	3.81	0.80	3.54	0.79	3.87	0.78
		Hawaiian or Native Pacific Islander	0						
		Hispanic or Latinx/Latine	14	3.73	0.78	3.58	0.95	3.87	0.76
		North African or Middle Eastern	5	3.98	0.85	4.05	0.77	4.30	0.77
		White	281	3.86	0.81	3.72	1.07	4.01	1.02

REFERENCES

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology,* 105(3), 399-412.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Marcoulides, K. M., & Yuan, K. H. (2017). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 148-153.

Peugh, J., & Feldon, D. F. (2020). "How well does your structural equation model fit your data?": Is Marcoulides and Yuan's equivalence test the answer? *CBE—Life Sciences Education*, *19*(3), es5.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software,* 48(2), 1–36.

Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 319-330.