

# PASS

Promoting At-Promise Student Success

## COMPILATION OF PSYCHOMETRIC METHODOLOGY AND ANALYSIS OF TSLC SCALES

**SAMANTHA NEIMAN, MARK MASTERTON, AND EVAN NIELSEN**  
**AMERICAN INSTITUTES FOR RESEARCH**

**DECEMBER 20, 2020**



# Memo

---

**Date:** December 22, 2020  
**To:** Thompson Scholars Learning Community (TSLC) Evaluation Project Team  
**From:** Samantha Neiman, Mark Masterton, and Evan Nielsen  
**Re:** Compilation of Psychometric Methodology and Analysis of TSLC Scales

---

## Contents

- Introduction .....3**
- Brief Summary of Methods .....3**
- Introduction to Psychometric Diagnostics .....4**
- Initial “Baseline” Survey Scales .....6**
  - 1. Self-Efficacy: Expected Social Adjustment [T0SESA] .....6
  - 2. Self-Efficacy: Expected Academic Adjustment [T0SEAA] ..... 10
  - 3. Expected Mattering: Campus [T0MATCAMP] ..... 14
  - 4. Expected Sense of Belonging: Campus [T0SOBCAMP] ..... 18
  - 5. Expected Mattering: Learning Community [T0MATLC].....22
  - 6. Resiliency (Cohort 2015 T0) [T0RESILIENCY].....26
  - 7. Resiliency (Cohort 2016 T0) [T0RESILIENCY].....29
  - 8. Expected Sense of Belonging: Learning Communities [T0SOB].....32
  - 9. Expected Adapted Perceived Academic Control [T0APAC].....36
  - 10. Peer Interaction, High School [T0PIHS].....39
  - 11. Faculty Interaction, High School [T0FIHS].....43
  - 12. Faculty Encouragement, High School [T0FEHS].....46
- Follow-Up Survey Scales.....49**
  - 13. Self-Efficacy: Social Adjustment [T#SESA] .....49
  - 14. Self-Efficacy: Academic Adjustment [T#SEAA] .....52
  - 15. Mattering: Campus [T#MATCAMP].....56
  - 16. Sense of Belonging: Campus [T#SOBCAMP] .....60
  - 17. Mattering: Learning Community [T#MATLC].....64
  - 18. Validation [T#VALID].....68
  - 19. Adapted Perceived Academic Control [T#APAC].....72
  - 20. Mentoring Support [T#MENTOR].....76

21. Engagement—First-Year Seminar Courses [T#ENGFYS].....	79
22. Career Decision Making Self-Efficacy [T#CDMSE].....	84
23. Academic Peer Interaction, Non Learning Community [T#APINLC].....	88
24. Academic Peer Interaction, Learning Community About Non Learning Community [T#APILCNLC] .....	91
25. Academic Peer Interaction, Learning Community [T#APILC] .....	94
26. Social Peer Interaction, Non Learning Community [T#SPINLC] .....	97
27. Social Peer Interaction, Learning Community About Non Learning Community [T#SPILCNLC] .....	100
28. Social Peer Interaction, Learning Community [T#SPILC] .....	103
29. Engagement—Shared Academic Courses [T#ENGSAC].....	106
30. Faculty Course-Related Interaction, Non Learning Community [T#FCRNLC].....	110
31. Faculty Course-Related Interaction, Learning Community About Non Learning Community [T#FCRLCNLC] .....	113
32. Faculty Course-Related Interaction, Learning Community [T#FCRLC].....	116
33. Faculty Noncourse-Related Interaction, Non Learning Community [T#FNCRNLC].....	119
34. Faculty Noncourse-Related Interaction, Learning Community About Non Learning Community [T#FNCRLCNLC] .....	122
35. Faculty Noncourse-Related Interaction, Learning Community [T#FNCRLC].....	125
36. Faculty Encouragement, Non Learning Community [T#FENLC].....	128
37. Faculty Encouragement, Learning Community About Non Learning Community [T#FELCNLC] .....	131
38. Faculty Encouragement, Learning Community [T#FELC].....	134
39. Educational Resiliency [T#RESILIENCY].....	137
40. College Knowledge [T#CK].....	141
41. Staff Care and Support [T#SCS] .....	144
42. Sense of Belonging: Learning Communities [T#SOB].....	148
43. Academic Self-Efficacy: Study Skills (T1) [T1ASES] .....	152
44. Peer Collaboration in Shared Academic Courses (T1) [T1SACP] .....	155
45. Academic Validation [T#AVAL].....	157
46. Interpersonal Validation [T#IVAL] .....	160
47. Financial Stress [*Not Delivered*] .....	163
<b>Methods to Scale Across Timepoints .....</b>	<b>166</b>
<b>References .....</b>	<b>172</b>

## Introduction

This memo presents descriptive information on the psychometric properties of the following scales, using data from the 2015 and 2016 fall “baseline” surveys and the 2016, 2017, and 2018 spring “follow-up” surveys of the Evaluation of the Thompson Scholars Learning Community (TSLC) Program. Each of these scales is comprised of a set of items that, together, are intended to measure a construct (or latent trait) about students’ experiences of their campus, learning community, and self-perceptions.

Scales 1-12 are from the initial, fall, “baseline” surveys, while scales 13-47 are from the spring, follow-up surveys. The heading for each scale (also shown in the contents above) includes the related variable name from the analysis data file; the hash symbol (e.g., T#ABCD) is a wildcard that indicates a digit 1-4 should replace the hash, corresponding to each follow-up timepoint, T1 through T4.

## Brief Summary of Methods

The main goal of the psychometric analyses was to create scale scores using the Rasch rating scale model that would allow for direct comparison between cohorts and timepoints (for example, to examine growth over time) for the TSLC student survey measures. The survey remained largely consistent across cohorts and timepoints, however, several measures had changes in response options (4-point versus 7-point) and items (items were added, dropped, or reworded across time). To account for these changes, the analysis team used psychometric anchoring and equating techniques and, in some cases, collapsed response scales on the back end of the data. Our detailed approach for equating each scale is described in the Psychometric Methods section at the end of this document.

## Introduction to Psychometric Diagnostics

Although constructs (e.g., “self-efficacy”) are not directly measurable, responses to the survey items are used to generate estimates of an individual’s score relative to a construct. To produce reliable estimates of person scores for a wide range of individuals, it is important to design surveys that have varying levels of item difficulty represented (i.e., where some items on the survey are easy to positively endorse, some are difficult to positively endorse, and some are in between).

Rasch measurement models (such as the Rasch rating scale model used in this analysis) (Andrich, 1978; Wright & Masters, 1982) enable us to transform ordinal responses (e.g., “Strongly disagree”) into interval scale measures, and to evaluate the psychometric functioning of the scales (i.e., Is this scale producing reliable and valid data for interpretation by analysts and stakeholders?). Unlike other scaling methods, such as item response theory, the Rasch rating scale model is not sample dependent (Granger, 2008) and is ideal for examination of measurement properties of a scale used beyond the scope of the data set being examined. Furthermore, the Rasch rating scale model can be used to create construct-level scores for individuals that summarize the results of their responses to multiple items that take into account the varying difficulty of endorsing the items that make up the scale.

Rasch analyses were conducted using WINSTEPS (Linacre, 2018) to evaluate the psychometric properties of each of the scales. In this memo we report the results of our examination of the validity of the scales, including data for fit to the Rasch rating scale model and reliability. Diagnostics that speak to aspects of validity and are presented in this memo include the following:

**Item Difficulties and Hierarchy.** Item difficulty measures, presented in logits, represent the relative position of an item regarding how easy or difficult it is for respondents to positively endorse an item. Logit values typically range from about -5 to +5 and are centered at zero, meaning that an item of average difficulty will have a logit value of zero, items that are easier to endorse than average will have a negative logit value, and items that are more difficult to endorse than average will have a positive logit value. When scales include a range of items with varying levels of difficulty, it becomes possible to distinguish among the performance of a wider range of individuals.

**Reliability Coefficients.** To assess the reliability (i.e., the extent to which scores are free from measurement error) of each scale, we present two reliability coefficients: Cronbach’s alpha and Rasch reliability (also referred to as “person separation reliability”). Rasch reliability indicates the extent to which the measure can distinguish among different ability levels (e.g., people who are “low,” “medium,” or “high” on the latent trait). Like Cronbach’s alpha, Rasch reliability ranges from 0 to 1, with higher values representing higher reliability. Rasch reliability is typically lower than Cronbach’s alpha and, unlike Cronbach’s alpha, is independent of sample size. Rasch reliability coefficients can broadly be interpreted as  $0.9 = 3$ , or four distinct ability levels;  $0.8 = 2$ , or three distinct ability levels; or  $0.5 = 1$ , or two distinct ability levels.

**Data Fit to Model and Dimensionality.** The Rasch model assumes that one latent trait is measured by the items making up the scale. To confirm that the data fit this Rasch model assumption, we present the observed versus expected percentage of variance explained by the measure and examine whether there are other large contrasts (factors) in the data that would

indicate potential issues with multidimensionality. That is, if factors other than the measure account for a substantial percentage of variance in the data, it may be an indication that the items are measuring two or more latent traits instead of the intended one latent trait.

**Item Fit.** Weighted<sup>1</sup> and unweighted<sup>2</sup> item fit statistics indicate the extent to which each item contributes a data signal to the overall measure (i.e., contributes meaningful information to the scale). Item fit statistics are presented as mean square (MNSQ) estimates, which are expected to be between 0.5 and 2, with a value of 1 indicating an ideal item fit. MNSQ estimates above 2 indicate that an item’s ratio of noise to signal is too high, and MNSQ estimates below 0.5 indicate that the signal provided by the item is too weak to contribute meaningfully to the measure.

**Average Person Ability by Response Category.** Respondents’ position relative to the latent trait is referred to as their “ability.” One way to check that each item and response category meaningfully corresponds to the underlying construct is to confirm that people with greater ability on the construct select a more positive response category to each item, and vice versa. When this expectation is not met, it may be an indication that the response categories are not properly ordered, distinct, or appropriate for the item, or that the item does not fit the construct.

**Rating Scale Thresholds.** In the Rasch rating scale model, Andrich thresholds between response categories represent the point (in person ability logits) at which the probability is equal that a person will respond to one of the two adjacent categories (e.g., the Andrich threshold between categories 1 and 2 is the point at which someone has an equal probability of responding to category 1 or category 2). Andrich thresholds are examined for each scale to determine the extent to which response categories are properly ordered and distinct.

**Differential Item Functioning (DIF).** Each item was examined by gender<sup>3</sup> and race/ethnicity<sup>4</sup> subgroups in a DIF analysis to determine whether there are items that appear to be easier or more difficult for a subgroup of respondents. The threshold for DIF used in this analysis was a DIF contrast that is both significant and more than 0.5 logits apart, meaning that the item was easier or more difficult to endorse for the focal group (e.g., females) by one half a logit, taking overall person ability into account. If evidence of DIF is present for an item, University of Southern California may want to consider further examination of the item to determine whether DIF is problematic for measurement purposes. For some scales identified as priority measures (Tier I), DIF also was examined by timepoint and cohort to confirm invariance in the functioning of items in each scale.

---

<sup>1</sup> Item **infit** is an inlier-pattern-sensitive fit statistic based on the chi-square statistic with each observation weighted by its statistical information (model variance). This is more sensitive to unexpected patterns of observations by persons on items that are roughly targeted on them. For more information on infit statistics, see <https://www.winsteps.com/winman/misfitdiagnosis.htm>.

<sup>2</sup> Item **outfit** is an outlier-sensitive fit statistic based on the conventional chi-square statistic and is more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them. For more information on outfit statistics, see <https://www.winsteps.com/winman/misfitdiagnosis.htm>.

<sup>3</sup> Gender categories examined in DIF analysis included male and female.

<sup>4</sup> Racial/ethnic categories examined in DIF analysis included Black or African American and non-Spanish-Hispanic-Latino (“Black”), Asian and non-Spanish-Hispanic-Latino (“Asian”), Spanish-Hispanic-Latino and Any Race (“Latino”), and White and non-Spanish-Hispanic-Latino (“White”).

## Initial “Baseline” Survey Scales

### 1. Self-Efficacy: Expected Social Adjustment [T0SESA]

#### *Items*

This scale consists of eight items that ask students about their expected social experiences at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Self-Efficacy: Expected Social Adjustment.

The response options for each item in this scale are:

Cannot do this at all ○ ○ ○ ○ ○ ○ ○ Absolutely can do this

#### Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0	Cohort 2016, T0
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:		
[L.1] Fit in socially at {INSTITUTION} <sup>1</sup>	✓	✓
[L.2] Get involved in interesting activities	✓	✓
[L.3] Adjust well to life as a college student <sup>2</sup>	✓	✓
[L.4] Develop good relationships with other students	✓	✓
[L.5] Feel at ease with others	✓	✓
[L.6] Meet people and make friends	✓	✓
[L.7] Make friends you can talk about your very personal problems with	X	✓
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}	X	✓

1. “at {INSTITUTION}” was added to the survey item between survey administrations.

2. “to life as a college student” was added to the survey item between survey administrations.

#### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.34 to 0.75.

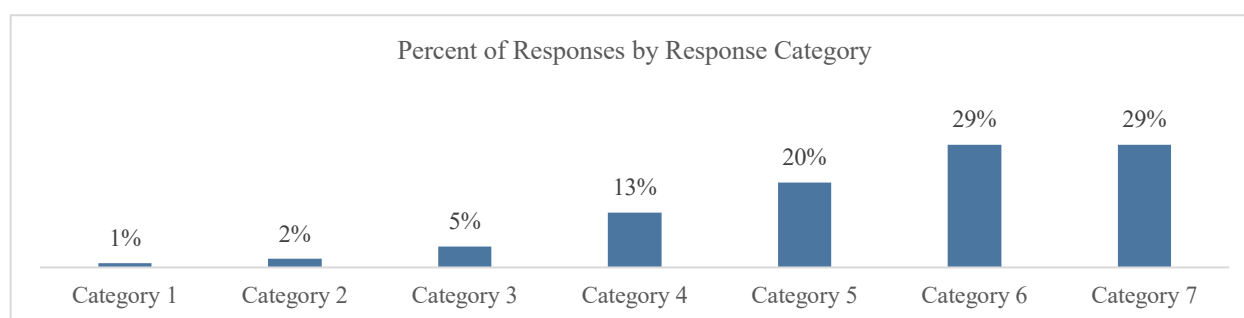
Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:	
[L.8] Be completely satisfied with your decision to attend {INSTITUTION} [ <i>easiest item to endorse</i> ]	-0.73
[L.4] Develop good relationships with other students	-0.29
[L.3] Adjust well to life as a college student	-0.14
[L.6] Meet people and make friends	-0.09
[L.1] Fit in socially at {INSTITUTION}	0.00
[L.5] Feel at ease with others	0.15
[L.2] Get involved in interesting activities	0.26

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:	
[L.7] Make friends you can talk about your very personal problems with <i>[most difficult item to endorse]</i>	0.96

### ***Data and Psychometric Diagnostics***

Of the 2,225 people who provided data for this scale, 137 had maximum extreme scores (i.e., selected the most positive response option for all items), and 1 had a minimum extreme score (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.94
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 59% of the variance in the data (of the 57.5% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:		
[L.1] Fit in socially at {INSTITUTION}	0.76	0.76
[L.2] Get involved in interesting activities	1.12	1.08
[L.3] Adjust well to life as a college student	1.06	1.04



Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:		
[L.4] Develop good relationships with other students	0.62	0.65
[L.5] Feel at ease with others	0.77	0.75
[L.6] Meet people and make friends	0.74	0.76
[L.7] Make friends you can talk about your very personal problems with	1.30	1.26
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}	1.63	1.67

### Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:							
[L.1] Fit in socially at {INSTITUTION}	-1.17	-0.67	-0.22	0.41	1.08	1.87	3.34
[L.2] Get involved in interesting activities	-1.05	-0.32	0.10	0.65	1.30	1.93	3.41
[L.3] Adjust well to life as a college student	-2.21	-0.79	-0.11	0.50	1.10	1.76	3.16
[L.4] Develop good relationships with other students	-1.69	-1.02	-0.40	0.17	0.85	1.66	3.24
[L.5] Feel at ease with others	-1.51	-0.56	-0.10	0.51	1.16	2.02	3.55
[L.6] Meet people and make friends	-1.33	-0.84	-0.09	0.34	1.08	1.74	3.21
[L.7] Make friends you can talk about your very personal problems with	-0.53	0.13	0.68	1.27	1.77	2.17	3.63
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}	-0.92	-0.49	0.41	<b>0.41</b>	0.86	1.42	2.51

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.20
2	-1.07
3	-0.62
4	0.36
5	1.14
6	2.39

### **Summary**

This scale functions well as a measure of Self-Efficacy: Expected Social Adjustment. There is evidence that the scale measures one underlying latent trait, reliably differentiates between

people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 2. Self-Efficacy: Expected Academic Adjustment [TOSEAA]

### Items

This scale consists of 15 items that ask students about their expectations of the kind of student they will be academically at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Self-Efficacy: Expected Academic Adjustment.

The response options for each item in this scale are:

Cannot do this at all ○ ○ ○ ○ ○ ○ ○ Absolutely can do this

### Changes Over Time:

Three items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0	Cohort 2016, T0
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:		
[M.1] Finish my assignments on time	✓	✓
[M.2] Get myself to study even when there are other interesting things to do	✓	✓
[M.3] Take good notes during class	✓	✓
[M.4] Organize my schoolwork	✓	✓
[M.5] Remember information presented in class	✓	✓
[M.6] Remember information presented in textbooks	✓	✓
[M.7] Arrange a place to study without distractions	✓	✓
[M.8] Motivate myself to do schoolwork	✓	✓
[M.9] Contribute to class discussions	✓	✓
[M.10] Understand what my professors expect of me academically	✓	✓
[M.11] Have effective study skills	✓	✓
[M.12] Adjust to the academic demands of college	✓	✓
[M.16] Work your hardest to do as well as you can	X	✓
[M.17] Be completely satisfied with your academic performance	X	✓
[M.18] Always keep up-to-date with your schoolwork	X	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.05 to 0.87.

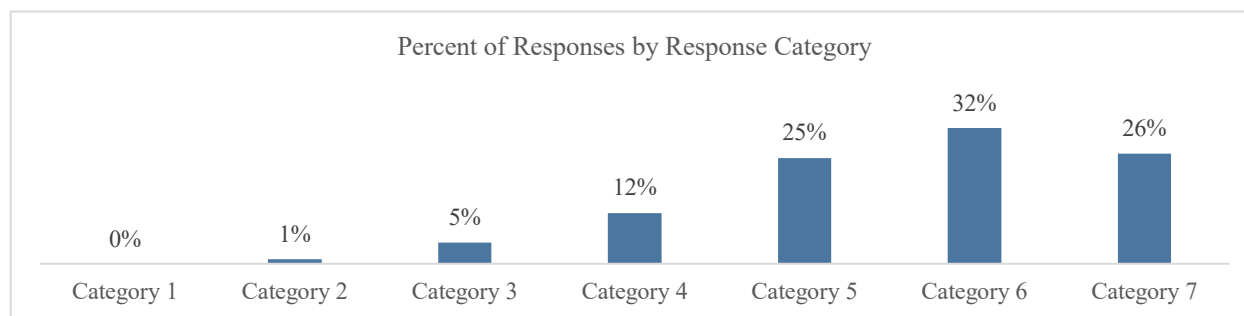
Item	Item Difficulty (logits)
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:	
[M.16] Work your hardest to do as well as you can [ <i>easiest item to endorse</i> ]	-1.05
[M.1] Finish my assignments on time	-0.59
[M.18] Always keep up-to-date with your schoolwork	-0.40
[M.4] Organize my schoolwork	-0.24
[M.10] Understand what my professors expect of me academically	-0.17

Item	Item Difficulty (logits)
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:	
[M.7] Arrange a place to study without distractions	-0.16
[M.3] Take good notes during class	-0.01
[M.17] Be completely satisfied with your academic performance	0.00
[M.12] Meet the academic demands of college	0.06
[M.8] Motivate myself to do schoolwork	0.08
[M.11] Have effective study skills	0.12
[M.2] Get myself to study even when there are other interesting things to do	0.27
[M.5] Remember information presented in class	0.49
[M.9] Contribute to class discussions	0.74
[M.6] Remember information presented in textbooks <i>[most difficult item to endorse]</i>	0.87

### ***Data and Psychometric Diagnostics***

Of the 2,225 people who provided data for this scale, 40 had maximum extreme scores (i.e., selected the most positive response option for all items), and 0 had the minimum extreme score (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### **Reliability Coefficients:**

- Cronbach's  $\alpha$  (inter-item consistency): 0.97
- Rasch reliability: 0.88

#### **Data Fit to Model and Dimensionality:**

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 50.8% of the variance in the data (of the 52.3% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:		
[M.1] Finish my assignments on time	1.03	1.09
[M.2] Get myself to study even when there are other interesting things to do	1.09	1.09
[M.3] Take good notes during class	1.16	1.18
[M.4] Organize my schoolwork	1.06	1.07
[M.5] Remember information presented in class	0.87	0.86
[M.6] Remember information presented in textbooks	1.03	1.02
[M.7] Arrange a place to study without distractions	1.22	1.24
[M.8] Motivate myself to do schoolwork	0.96	0.97
[M.9] Contribute to class discussions	1.64	1.63
[M.10] Understand what my professors expect of me academically	0.89	0.90
[M.11] Have effective study skills	0.67	0.66
[M.12] Meet the academic demands of college	0.77	0.76
[M.16] Work your hardest to do as well as you can	0.91	1.03
[M.17] Be completely satisfied with your academic performance	1.18	1.08
[M.18] Always keep up-to-date with your schoolwork	0.71	0.73

Ability Mean by Category:

With the exceptions noted in red, the data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:							
[M.1] Finish my assignments on time	0.14*	-0.57	0.06	0.56	1.12	1.74	2.91
[M.2] Get myself to study even when there are other interesting things to do	-0.48	-0.08	0.56	0.97	1.59	2.16	3.50
[M.3] Take good notes during class	-0.91	0.06	0.50	0.86	1.39	1.98	3.24
[M.4] Organize my schoolwork	1.28	-0.07	0.34	0.62	1.30	1.94	3.11
[M.5] Remember information presented in class	-1.22	0.01	0.45	0.97	1.56	2.36	3.87
[M.6] Remember information presented in textbooks	-0.08	0.51	0.68	1.15	1.80	2.58	4.14
[M.7] Arrange a place to study without distractions	-0.54	-0.16	0.43	0.85	1.34	2.00	3.11
[M.8] Motivate myself to do schoolwork	-0.30	-0.02	0.40	0.83	1.33	2.06	3.46
[M.9] Contribute to class discussions	0.35	0.69	0.88	1.28	1.73	2.23	3.40

Item Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how certain you are that you will be able to do the following:	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[M.10] Understand what my professors expect of me academically	#	0.17*	0.08	0.61	1.26	1.93	3.30
[M.11] Have effective study skills	-0.55*	-0.93	0.23	0.66	1.36	2.13	3.66
[M.12] Meet the academic demands of college	-1.00	-0.49	0.19	0.69	1.31	2.13	3.49
[M.16] Work your hardest to do as well as you can	-2.23	-0.52	-0.24	0.29	0.93	1.64	2.76
[M.17] Be completely satisfied with your academic performance	-0.56	0.54	0.76	0.71	1.36	2.08	3.41
[M.18] Always keep up-to-date with your schoolwork	-1.45	-0.80	-0.09	0.40	1.15	1.87	3.27

\*Based on  $n$  size less than 12. Interpret with caution.

#No data.

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options (i.e., respondents do not necessarily see each as distinct). Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.73
2	-1.26
3	-0.56
4	0.31
5	1.42
6	2.82

### **Summary**

This scale functions well as a measure of Self-Efficacy: Expected Academic Adjustment. There is evidence that the scale measures one underlying latent trait, reliably differentiates between people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

### 3. Expected Mattering: Campus [T0MATCAMP]

#### *Items*

This scale consists of eight items that ask students about supportive relationships they expect to have at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Expected Mattering: Campus.

The response options for each item in this scale are<sup>5</sup>:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

#### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0 (4-point ROs)	Cohort 2016, T0 (7-point ROs)
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[O.1] There will be people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	✓	✓
[O.3] There will be people at {INSTITUTION} who are generally supportive of my individual needs. <sup>1</sup>	✓	✓
[O.4] There will be people at {INSTITUTION} who seem happy about my accomplishments. <sup>1</sup>	✓	✓
[O.6] There will be people at {INSTITUTION} who are concerned about my future.	✓	✓
[O.7] There will be people at {INSTITUTION} who are genuinely interested in me as a person.	✓	✓
[O.8] Other students at {INSTITUTION} will be happy for me when I do well on exams or projects. <sup>2</sup>	✓	✓
[O.15] People I value at {INSTITUTION} will be disappointed when I don't accomplish all I should.	X	✓
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	X	✓

<sup>1</sup> “that” was changed to “who” between survey administrations.

<sup>2</sup> “at {INSTITUTION}” was added to the survey item between survey administrations.

#### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.72 to 0.83.

<sup>5</sup> Students in Cohort 2015 were presented with four response options. To be scaled together, data from the Cohort 2016 survey were collapsed and scale diagnostics were examined with 4-point response options. See the Methods to Scale Across Timepoints section for more information.

Item Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	Item Difficulty (logits)
[O.4] There will be people at {INSTITUTION} who seem happy about my accomplishments. <i>[easiest item to endorse]</i>	-0.72
[O.3] There will be people at {INSTITUTION} who are generally supportive of my individual needs. <sup>2</sup>	-0.46
[O.6] There will be people at {INSTITUTION} who are concerned about my future.	-0.36
[O.7] There will be people at {INSTITUTION} who are genuinely interested in me as a person.	-0.23
[O.8] Other students at {INSTITUTION} will be happy for me when I do well on exams or projects.	0.06
[O.15] People I value at {INSTITUTION} will be disappointed when I don't accomplish all I should.	0.16
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	0.73
[O.1] There will be people at {INSTITUTION} who are sad for me when I fail in something I set out to do. <i>[most difficult item to endorse]</i>	0.83

### ***Data and Psychometric Diagnostics***

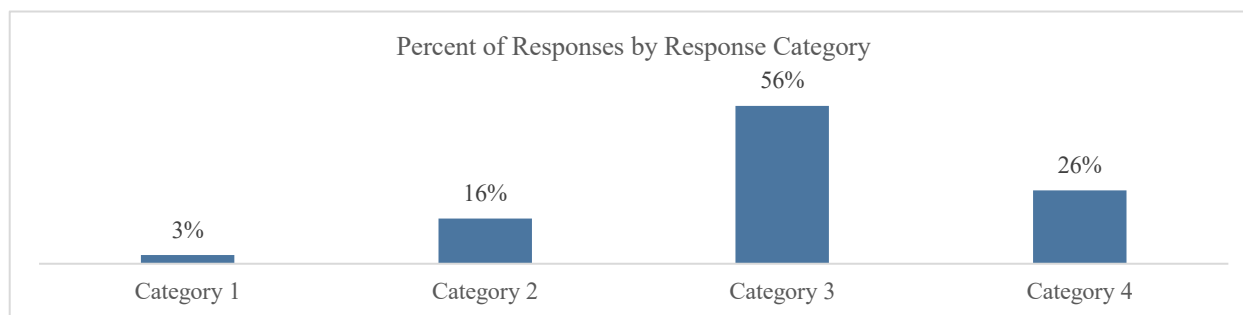
Of the 2,225 people who provided data for this scale, 120 had maximum extreme scores (i.e., selected the most positive response option for all items), and five had the minimum extreme score (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4



Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.90
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 50.6% of the variance in the data (of the 49.0% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[O.1] There will be people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	1.06	1.04
[O.3] There will be people at {INSTITUTION} who are generally supportive of my individual needs. <sup>2</sup>	0.68	0.69
[O.4] There will be people at {INSTITUTION} who seem happy about my accomplishments.	0.64	0.68
[O.6] There will be people at {INSTITUTION} who are concerned about my future.	0.95	0.97
[O.7] There will be people at {INSTITUTION} who are genuinely interested in me as a person.	0.79	0.80
[O.8] Other students at {INSTITUTION} will be happy for me when I do well on exams or projects.	0.94	0.94
[O.15] People I value at {INSTITUTION} will be disappointed when I don't accomplish all I should.	0.95	0.97
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	1.58	1.51

Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
[O.1] There will be people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	-1.25	0.14	1.75	4.17
[O.3] There will be people at {INSTITUTION} who are generally supportive of my individual needs.	-3.05	-0.77	1.23	3.88
[O.4] There will be people at {INSTITUTION} who seem happy about my accomplishments.	-3.47	-0.87	1.07	3.73
[O.6] There will be people at {INSTITUTION} who are concerned about my future.	-1.53	-0.32	1.17	3.83
[O.7] There will be people at {INSTITUTION} who are genuinely interested in me as a person.	-2.46	-0.48	1.32	3.90
[O.8] Other students at {INSTITUTION} will be happy for me when I do well on exams or projects.	-1.93	-0.06	1.52	4.16
[O.15] People I value at {INSTITUTION} will be disappointed when I don't accomplish all I should.	-1.64	-0.34	1.37	4.13
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	-0.47	0.15	1.57	3.96

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.72
2	-0.53
3	3.26

**Summary**

This scale functions well as a measure of Expected Mattering: Campus. There is evidence that the scale measures one underlying latent trait, reliably differentiates between people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 4. Expected Sense of Belonging: Campus [TOSOB CAMP]

### Items

This scale consists of eight items that ask students about their expected sense of belonging at their institution. Five of the eight items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and three of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of sense of belonging.

The response options for each item in this scale are<sup>6</sup>:

Strongly Disagree                                Strongly Agree

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0 (4-point ROs)	Cohort 2016, T0 (7-point ROs)
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[Q.1] I will feel like an outsider.*	✓	✓
[Q.2] I will make friends easily.	✓	✓
[Q.3] I will feel like I belong.	✓	✓
[Q.4] I will feel awkward and out of place.*	✓	✓
[Q.5] I will feel lonely.*	✓	✓
[Q.6] I expect other students will like me.	✓	✓
[Q.7] I will see myself as an important part of the {INSTITUTION} community.	X	✓
[Q.8] I will feel I am a member of the {INSTITUTION} community.	X	✓

\*Negatively valenced item; reverse coded.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.69 to 0.53.

Item	Item Difficulty (logits)
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[Q.8] I will feel I am a member of the {INSTITUTION} community. <i>[easiest item to endorse]</i>	-0.69
[Q.3] I will feel like I belong.	-0.34
[Q.6] I expect other students will like me.	-0.29

<sup>6</sup> Students in Cohort 2015 were presented with four response options in timepoint 0. To be scaled together, data from subsequent administrations were collapsed and scale diagnostics were examined with 4-point response options. See the Methods to Scale Across Timepoints section for more information.

Item	Item Difficulty (logits)
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[Q.7] I will see myself as an important part of the {INSTITUTION} community.	-0.18
[Q.2] I will make friends easily.	-0.11
[Q.5] I will feel lonely.*	0.19
[Q.1] I will feel like an outsider.*	0.31
[Q.4] I will feel awkward and out of place.* <i>[most difficult item to endorse]</i>	0.53

\*Negatively valenced item; reverse coded.

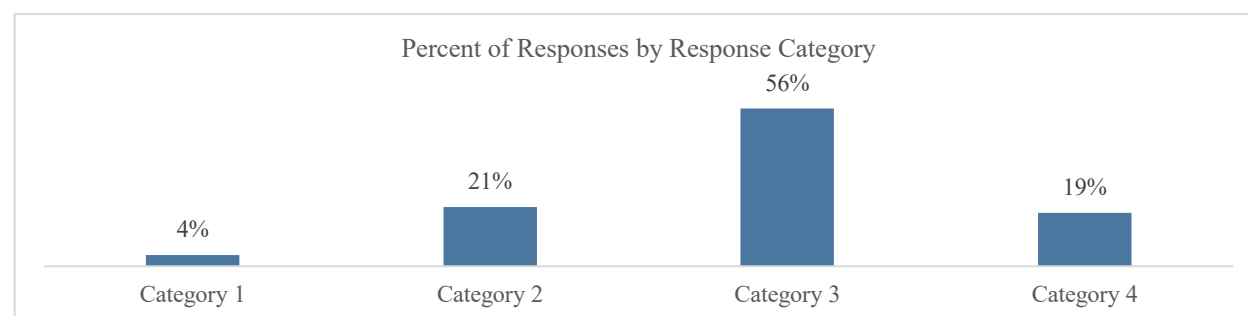
### ***Data and Psychometric Diagnostics***

Of the 2,225 people who provided data for this scale, 76 had maximum extreme scores (i.e., selected the most positive response option for all items), and 6 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Across all items in this scale, the percentage of responses in each category is as follows:



### **Reliability Coefficients:**

- Cronbach's  $\alpha$  (inter-item consistency): 0.89
- Rasch reliability: 0.82

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 51.8% of the variance in the data (of the 49.3% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for 11.4% of unexplained variance, a slightly notable percentage but substantially lower than variance accounted for by the measure. Not uncommon in scales with positively and negatively valenced items, the largest contrast is defined by the positively valenced and negatively valenced items in this scale. There are no other large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[Q.1] I will feel like an outsider.*	0.88	0.89
[Q.2] I will make friends easily.	0.85	0.87
[Q.3] I will feel like I belong.	0.68	0.70
[Q.4] I will feel awkward and out of place.*	0.83	0.82
[Q.5] I will feel lonely.*	0.99	0.97
[Q.6] I expect other students will like me.	1.05	1.04
[Q.7] I will see myself as an important part of the {INSTITUTION} community.	1.08	1.07
[Q.8] I will feel I am a member of the {INSTITUTION} community.	0.97	0.99

\*Negatively valenced item; reverse coded.

Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:				
[Q.1] I feel like an outsider.*	-1.41	-0.27	1.29	3.60
[Q.2] I make friends easily.	-2.15	-0.53	1.18	3.51
[Q.3] I feel like I belong.	-2.79	-0.78	1.05	3.52
[Q.4] I feel awkward and out of place.*	-1.54	-0.16	1.40	3.72
[Q.5] I feel lonely.*	-1.60	-0.19	1.24	3.41
[Q.6] I believe other students like me.	-1.45	-0.32	1.35	3.65
[Q.7] I see myself as an important part of the {INSTITUTION} community.	-2.31	-0.54	0.96	3.19
[Q.8] I feel I am a member of the {INSTITUTION} community.	-2.91	-0.87	0.75	2.93

\*Negatively valenced item; reverse coded.

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

	Positively Valenced Response Scale	Negatively Valenced Response Scale
Category	Threshold	Threshold
1	-2.85	-2.25
2	-0.39	-0.33
3	3.24	2.58

***Summary***

This scale functions well as a measure of Expected Sense of Belonging: Campus. There is evidence that the scale measures one underlying latent trait, despite a small contrast attributed to the mix of positively and negatively valenced items. Furthermore, there is evidence that this scale reliably differentiates between people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.

## 5. Expected Mattering: Learning Community [TOMATLC]

### Items

This scale consists of seven items that ask students about supportive relationships they expect to have at their learning community.<sup>7</sup> All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Expected Mattering: Learning Community.

The response options for each item in this scale are:

Strongly Disagree ○ ○ ○ ○ ○ ○ ○ Strongly Agree

### Item Changes Over Time:

This scale was included in the 2016 cohort baseline survey only; therefore, there were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.57 to 0.54.

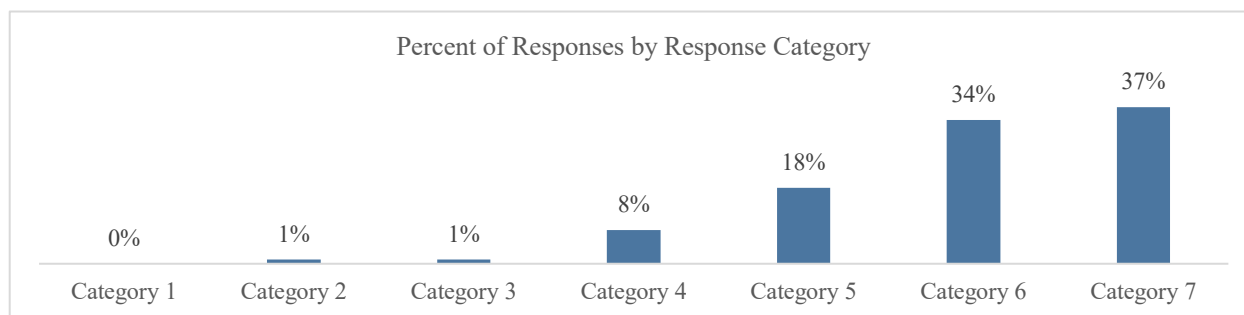
Item	Item Difficulty (logits)
Thinking only about the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[O.11] There will be people at the {COMMUNITY} who seem happy about my accomplishments. <i>[easiest item to endorse]</i>	-0.57
[O.10] There will be people at the {COMMUNITY} who are generally supportive of my individual needs.	-0.34
[O.12] There will be people at the {COMMUNITY} who are concerned about my future.	-0.24
[O.13] There will be people at the {COMMUNITY} who are genuinely interested in me as a person.	-0.11
[O.14] Other students at the {INSTITUTION} will be happy for me when I do well on exams or projects.	0.31
[O.9] There will be people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	0.41
[O.17] People I value at the {INSTITUTION} are disappointed when I don't accomplish all I should. <i>[most difficult item to endorse]</i>	0.54

### Data and Psychometric Diagnostics

Of the 526 people who provided data for this scale, 89 had maximum extreme scores (i.e., selected the most positive response option for all items), and 0 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

<sup>7</sup> One item (O.18. I feel pressure to do my best because people I value at the {INSTITUTION} otherwise would be disappointed) was found to be misfitting, with an outfit MNSQ of 2.06, and was dropped from the scale.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1, 2, and 3 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.94
- Rasch reliability: 0.80

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 56.1% of the variance in the data (of the 56.7% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for 10.5% of unexplained variance, a slightly notable percentage but substantially lower than variance accounted for by the measure.<sup>8</sup>

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking only about the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[O.9] There will be people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	1.36	1.29
[O.10] There will be people at the {COMMUNITY} who are generally supportive of my individual needs.	0.73	0.74
[O.11] There will be people at the {COMMUNITY} who seem happy about my accomplishments.	0.75	0.76
[O.12] There will be people at the {COMMUNITY} who are concerned about my future.	1.09	1.15
[O.13] There will be people at the {COMMUNITY} who are genuinely interested in me as a person.	0.82	0.79

<sup>8</sup> During our validation study of the follow-up survey, a similar notable contrast was found and further explored. Specifically, we examined an alternative version of this measure with items O.10, O.11, O.13, and O.14 (based on diagnostics for the full scale) and did not find notable improvements regarding dimensionality; therefore, we did not recommend changes.



Item	Outfit MNSQ	Infit MNSQ
Thinking only about the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[O.14] Other students at the {COMMUNITY} will be happy for me when I do well on exams or projects.	1.08	1.01
[O.17] People I value at the {COMMUNITY} are disappointed when I don't accomplish all I should.	1.34	1.22

### Ability Mean by Category:

With the exception of the values noted in red below, the data for this survey follow the expected pattern of average person ability measures by response categories. All of the exceptions noted in red are based on small sample sizes ( $n < 10$ ) and should therefore be interpreted with caution.

Average person ability measures by response category are as follows:

Item Thinking only about the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[O.9] There will be people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	-0.02	0.88	0.45	0.80	1.51	2.87	5.59
[O.10] There will be people at the {COMMUNITY} who are generally supportive of my individual needs.	#	-0.46	-1.31	0.20	1.07	2.51	5.25
[O.11] There will be people at the {COMMUNITY} who seem happy about my accomplishments.	#	-2.71	-1.63	0.10	0.93	2.41	5.01
[O.12] There will be people at the {COMMUNITY} who are concerned about my future.	#	0.15	0.41	0.31	1.17	2.43	5.22
[O.13] There will be people at the {COMMUNITY} who are genuinely interested in me as a person.	-2.71	-1.19	-0.09	0.25	1.41	2.54	5.44
[O.14] Other students at the {INSTITUTION} will be happy for me when I do well on exams or projects.	-1.43	1.54	0.13	0.81	1.48	2.79	5.65
[O.17] People I value at the {INSTITUTION} are disappointed when I don't accomplish all I should.	-0.83	-0.58	1.15	1.03	1.73	2.88	5.77

#No data.

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are disordered, with no peak for category 3. In other words, there is no point in the ability distribution where response category 3 is the highest probability response. This finding, combined with the thresholds' close distance ( $< 1$  logit) for some categories, indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.51
2	-1.25
3	-1.64
4	0.15
5	1.40
6	3.85

### ***Summary***

This scale functions fairly well as a measure of Expected Mattering: Learning Community. As is, there is evidence that this scale reliably differentiates between people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.

## 6. Resiliency (Cohort 2015 T0) [T0RESILIENCY]

### *Items*

This scale consists of ten items that ask students about their capacity to be resilient to challenging circumstances. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Resiliency.

The response options for each item in this scale are:

- Strongly agree
- Agree
- Disagree
- Strongly disagree

### Changes Over Time:

As a baseline scale, these questions were asked differently to cohort 2015 and cohort 2016. See the next section for the 2016 scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.93 to 1.18.

Item	Item Difficulty (logits)
Please indicate the extent to which you agree or disagree with the following statements ...	
[V6] I can achieve goals despite obstacles.	-0.93
[V9] I think of myself as a strong person.	-0.82
[V1] I am able to adapt to change.	-0.59
[V2] I can deal with whatever comes.	-0.27
[V3] I try to see the humorous side of problems.	0.05
[V10] I can handle unpleasant feelings.	0.14
[V7] I can stay focused under pressure.	0.29
[V5] I tend to bounce back after illness or hardship.	0.33
[V4] Coping with stress can strengthen me.	0.63
[V8] I am not easily discouraged by failure.	1.18

### *Data and Psychometric Diagnostics*

Of the 1,059 people who provided data for this scale, 29 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 1%
- Category 2 = 12%
- Category 3 = 57%

- Category 4 = 29%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.84
- Rasch reliability: 0.80

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 40.8% of the variance in the data (of the 40.7% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item Please indicate the extent to which you agree or disagree with the following statements ...	Outfit MNSQ	Infit MNSQ
[V1] I am able to adapt to change.	0.89	0.83
[V2] I can deal with whatever comes.	0.81	0.79
[V3] I try to see the humorous side of problems.	1.26	1.23
[V4] Coping with stress can strengthen me.	1.16	1.15
[V5] I tend to bounce back after illness or hardship.	1.03	1.03
[V6] I can achieve goals despite obstacles.	0.63	0.66
[V7] I can stay focused under pressure.	0.97	0.98
[V8] I am not easily discouraged by failure.	1.17	1.10
[V9] I think of myself as a strong person.	1.04	1.09
[V10] I can handle unpleasant feelings.	1.03	1.03

#### Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item Please indicate the extent to which you agree or disagree with the following statements ...	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
[V1] I am able to adapt to change.	-2.56	0.27	1.43	3.50
[V2] I can deal with whatever comes.	-2.90	0.32	1.52	3.73
[V3] I try to see the humorous side of problems.	-0.04	0.64	1.71	3.49
[V4] Coping with stress can strengthen me.	0.30	0.86	1.79	4.07
[V5] I tend to bounce back after illness or hardship.	0.14	0.72	1.65	4.00
[V6] I can achieve goals despite obstacles.	-2.87	-0.48	1.23	3.50
[V7] I can stay focused under pressure.	-0.42	0.56	1.67	3.97

Item Please indicate the extent to which you agree or disagree with the following statements ...	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
[V8] I am not easily discouraged by failure.	0.27	0.93	2.04	4.25
[V9] I think of myself as a strong person.	-1.32	0.31	1.33	3.27
[V10] I can handle unpleasant feelings.	-1.02	0.69	1.68	3.75

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.69
2	-0.59
3	3.29

### Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

### ***Summary***

This scale functions well as a measure of Resiliency. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 7. Resiliency (Cohort 2016 T0) [T0RESILIENCY]

### *Items*

This scale consists of seven items that ask students about their capacity to be resilient to challenging educational circumstances. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Resiliency.

The response options for each item in this scale are:

- Strongly agree
- Agree
- Disagree
- Strongly disagree

### Changes Over Time:

As a baseline scale, these questions were asked differently to cohort 2015 and cohort 2016. See the previous section for the 2015 scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.41 to 1.78.

Item	Item Difficulty (logits)
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements ...	
[V6] I can achieve my educational goals despite obstacles.	-1.41
[V1] I am able to adapt to change in educational situations or settings.	-0.66
[V2] I can deal with whatever comes in educational situations or settings.	-0.21
[V4] Coping with stress can strengthen me in educational challenges.	-0.14
[V7] I can stay focused on my assignments and coursework under pressure.	-0.04
[V10] I can handle unpleasant feelings related to my educational experiences.	0.68
[V8] I am not easily discouraged by failure in my courses.	1.78

### *Data and Psychometric Diagnostics*

Of the 1,166 people who provided data for this scale, 54 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 2%
- Category 2 = 13%
- Category 3 = 60%
- Category 4 = 26%

Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.78
- Rasch reliability: 0.75

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 46.1% of the variance in the data (of the 45.7% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements ...		
[V1] I am able to adapt to change in educational situations or settings.	0.89	0.89
[V2] I can deal with whatever comes in educational situations or settings.	0.68	0.71
[V4] Coping with stress can strengthen me in educational challenges.	1.29	1.31
[V6] I can achieve my educational goals despite obstacles.	0.85	0.84
[V7] I can stay focused on my assignments and coursework under pressure.	0.94	0.96
[V8] I am not easily discouraged by failure in my courses.	1.41	1.26
[V10] I can handle unpleasant feelings related to my educational experiences.	0.85	0.85

Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements ...				
[V1] I am able to adapt to change.	0.58	-.16*	1.19	3.93
[V2] I can deal with whatever comes.	-0.09	-.28*	1.36	4.39
[V3] I try to see the humorous side of problems.	-1.02	0.56	1.39	3.80
[V4] Coping with stress can strengthen me.	-1.45	-0.30	0.95	3.41
[V5] I tend to bounce back after illness or hardship.	-1.14	0.09	1.44	4.13
[V6] I can achieve goals despite obstacles.	-0.01	1.01	2.08	4.76
[V7] I can stay focused under pressure.	-0.38	0.33	1.70	4.92

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.66
2	-0.91
3	3.57

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions well as a measure of Resiliency. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.



## 8. Expected Sense of Belonging: Learning Communities [T0SOB]

### Items

This scale consists of eight items that ask students about their expected sense of belonging in their learning communities. Five of the eight items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and three of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of sense of belonging.

The response options for each item in this scale are<sup>9</sup>:

Strongly Disagree                                Strongly Agree

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0 (4-point ROs)	Cohort 2016, T0 (7-point ROs)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[R.1] I will feel like an outsider.*	✓	✓
[R.2] I will make friends easily.	✓	✓
[R.3] I will feel like I belong.	✓	✓
[R.4] I will feel awkward and out of place.*	✓	✓
[R.5] I will feel lonely.*	✓	✓
[R.6] I expect other students will like me.	✓	✓
[R.7] I will see myself as an important part of the {COMMUNITY} community.	X	✓
[R.8] I will feel I am a member of the {COMMUNITY} community.	X	✓

\*Negatively valenced item; reverse coded.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.23 to 0.79.

Item	Item Difficulty (logits)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[R.8] I will feel I am a member of the {COMMUNITY} community.	-1.23
[R.7] I will see myself as an important part of the {COMMUNITY} community.	-0.58
[R.3] I will feel like I belong.	-0.35

<sup>9</sup> Students in Cohort 2015 were presented with four response options in timepoint 0. To be scaled together, data from subsequent administrations were collapsed and scale diagnostics were examined with 4-point response options. See the Methods to Scale Across Timepoints section for more information.

Item	Item Difficulty (logits)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[R.2] I will make friends easily.	0.12
[R.5] I will feel lonely.*	0.21
[R.1] I will feel like an outsider.*	0.37
[R.4] I will feel awkward and out of place.*	0.66
[R.6] I expect other students will like me.	0.79

\*Negatively valenced item; reverse coded.

### ***Data and Psychometric Diagnostics***

Of the 1,059 people who provided data for this scale, 87 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 2%
- Category 2 = 14%
- Category 3 = 54%
- Category 4 = 30%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.79

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 49.8% of the variance in the data (of the 46.7% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for 12.9% of unexplained variance, a slightly notable percentage but substantially lower than variance accounted for by the measure. Not uncommon in scales with positively and negatively valenced items, the largest contrast is

defined by the positively valenced and negatively valenced items in this scale. There are no other large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[R.1] I will feel like an outsider.*	0.82	0.83
[R.2] I will make friends easily.	0.82	0.85
[R.3] I will feel like I belong.	0.66	0.69
[R.4] I will feel awkward and out of place.*	0.74	0.74
[R.5] I will feel lonely.*	0.90	0.88
[R.6] I expect other students will like me.	1.08	1.07
[R.7] I will see myself as an important part of the {COMMUNITY} community.	1.09	1.07
[R.8] I will feel I am a member of the {COMMUNITY} community.	1.06	1.03

\*Negatively valenced item; reverse coded.

#### Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:				
[R.1] I will feel like an outsider.*	-0.92	0.12	1.75	4.37
[R.2] I will make friends easily.	-1.87	-0.06	1.80	4.34
[R.3] I will feel like I belong.	-1.81	-0.63	1.56	4.26
[R.4] I will feel awkward and out of place.*	-0.71	0.08	1.81	4.60
[R.5] I will feel lonely.*	-0.83	0.09	1.71	4.31
[R.6] I expect other students will like me.	-0.06	0.42	2.03	4.82
[R.7] I will see myself as an important part of the {COMMUNITY} community.	-2.41	-0.54	1.45	3.63
[R.8] I will feel I am a member of the {COMMUNITY} community.	-5.29	-0.63	1.08	3.38

\*Negatively valenced item; reverse coded.

#### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

	<b>Positively Valenced Response Scale</b>	<b>Negatively Valenced Response Scale</b>
<b>Category</b>	<b>Threshold</b>	<b>Threshold</b>
1	-3.05	-2.42
2	-0.39	-0.35
3	3.44	2.76

### ***Summary***

This scale functions well as a measure of Expected Sense of Belonging: Learning Community. There is evidence that the scale measures one underlying latent trait, despite a small contrast attributed to the mix of positively and negatively valenced items. Furthermore, there is evidence that this scale reliably differentiates between people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.

## 9. Expected Adapted Perceived Academic Control [T0APAC]

### *Items*

This scale consists of four items that ask students about their capacity for control over their academic circumstances. Two of the four items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and two of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of Academic Control.

The response options for each item in this scale are:

- Strongly agree
- Agree
- Disagree
- Strongly disagree

### Changes Over Time:

There were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.99 to 0.89.

Item	Item Difficulty (logits)
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how strongly you agree with the following statements ...	
[S.2] The more effort I put into school, the better I will do.	-1.99
[S.1] I will have a great deal of control over my academic performance.	0.22
[S.4] There will be little I can do about my performance in school.	0.88
[S.3] No matter what I do, I will not do well in school.	0.89

### *Data and Psychometric Diagnostics*

Of the 2,225 people who provided data for this scale, 633 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 1%
- Category 2 = 3%
- Category 3 = 37%
- Category 4 = 60%

Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.64
- Rasch reliability: 0.40

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 41.2% of the variance in the data (of the 40.2% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for 25.6% of unexplained variance, a notable percentage that indicates possible multidimensionality, not uncommon in scales with positively and negatively valenced items, the largest contrast is defined by the positively valenced and negatively valenced items in this scale. There are no other large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how strongly you agree with the following statements ...		
[S.1] I will have a great deal of control over my academic performance.	1.07	1.09
[S.2] The more effort I put into school, the better I will do.	1.05	1.08
[S.3] No matter what I do, I will not do well in school.	1.01	0.97
[S.4] There will be little I can do about my performance in school.	0.86	0.86

Ability Mean by Category:

The data for this scale do not follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about the kind of student you expect you will be at {INSTITUTION}, please rate how strongly you agree with the following statements ...				
[S.1] I will have a great deal of control over my academic performance.	0.27	0.47	2.57	4.92
[S.2] The more effort I put into school, the better I will do.	0.14	-0.06*	1.72	4.27
[S.3] No matter what I do, I will not do well in school.	0.61	0.60*	2.39	4.90
[S.4] There will be little I can do about my performance in school.	0.26	0.82	2.35	4.94

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) in the response scale for the positively valenced items. However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. For the negatively valenced response scale, thresholds are disordered, with no peak for category 2 ("Disagree"); this

indicates that the response scale for these items may not be distinct and progress in the intended order for respondents. Andrich thresholds for this scale are as follows:

	<b>Positively Valenced Response Scale</b>	<b>Negatively Valenced Response Scale</b>
<b>Category</b>	<b>Threshold</b>	<b>Threshold</b>
1	-2.12	-1.10
2	-1.65	-1.52
3	3.76	2.62

### ***Summary***

This scale functions adequately as a measure of Expected Adapted Perceived Academic Control. There is evidence that the items all contribute meaningfully to the measure, but respondents do not use the rating scale as intended for the negatively valenced items, and potential issues of multidimensionality should be further explored.

## 10. Peer Interaction, High School [T0PIHS]

### Items

This scale consists of thirteen items that ask students about their academic interaction with peers during the final year of high school. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Five items were added to this measure (and 2 removed/collapsed) after the first survey administration as shown:

Item	Cohort 2015, T0	Cohort 2016, T0
During your final year in high school, how often did you do the following with fellow students...		
[C1] Shared your concerns about a class, test, or assignment?	✓	✓
[C2] Discussed something you learned in class?	✓	✓
[C3] Had a discussion with someone whose political opinions were very different than your own?	✓	X
[C4] Had a discussion with someone with very different religious beliefs than your own?	✓	X
[C5] Talked about social issues such as peace, justice, or human rights?	✓	✓
[C6] Discussed views about multiculturalism or diversity?	✓	✓
[C7] Had discussions with students whose personal values, [religious beliefs, or political opinions]* differed from your own?	✓	✓
[C8] Received advice about an academic issue?	✓	✓
[C9] Worked on a group project with a class mate?	✓	✓
[C10] Studied with another student?	✓	✓
[C11] Attended a social or cultural event with another student?	X	✓
[C12] Helped another student with a personal problem they were having?	X	✓
[C13] Helped another student by tutoring or teaching them?	X	✓
[C14] Talked about news or current events?	X	✓
[C15] Asked for advice about a personal concern or issue?	X	✓

\* In item C7, wording within brackets added for Cohort 2016, which replaced items C3 and C4.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.44 to 0.42.

Item	Item Difficulty (logits)
During your final year in high school, how often did you do the following with fellow students...	
[C2] Discussed something you learned in class?	-0.44
[C12] Helped another student with a personal problem they were having?	-0.40



Item	Item Difficulty (logits)
During your final year in high school, how often did you do the following with fellow students...	
[C14] Talked about news or current events?	-0.40
[C9] Worked on a group project with a class mate?	-0.35
[C1] Shared your concerns about a class, test, or assignment?	-0.25
[C5] Talked about social issues such as peace, justice, or human rights?	0.00
[C8] Received advice about an academic issue?	0.02
[C7] Had discussions with students whose personal values, [religious beliefs, or political opinions]* differed from your own?	0.06
[C10] Studied with another student?	0.26
[C6] Discussed views about multiculturalism or diversity?	0.34
[C15] Asked for advice about a personal concern or issue?	0.35
[C13] Helped another student by tutoring or teaching them?	0.39
[C11] Attended a social or cultural event with another student?	0.42

### ***Data and Psychometric Diagnostics***

Of the 2,225 people who provided data for this scale, 8 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 6%
- Category 2 = 15%
- Category 3 = 32%
- Category 4 = 29%
- Category 5 = 18%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.74

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 34.3% of the variance in the data (of the 33.6% of variance explained that would be expected if data are a good fit to the model). There is one small contrast accounting for 11.5% of unexplained variance.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
During your final year in high school, how often did you do the following with fellow students...		
[C1] Shared your concerns about a class, test, or assignment?	0.99	0.95
[C2] Discussed something you learned in class?	0.74	0.73
[C3] Had a discussion with someone whose political opinions were very different than your own?	0.95	0.93
[C4] Had a discussion with someone with very different religious beliefs than your own?	0.96	0.96
[C5] Talked about social issues such as peace, justice, or human rights?	0.95	0.95
[C6] Discussed views about multiculturalism or diversity?	0.85	0.83
[C7] Had discussions with students whose personal values, [religious beliefs, or political opinions]* differed from your own?	0.92	0.89
[C8] Received advice about an academic issue?	1.25	1.25
[C9] Worked on a group project with a class mate?	1.30	1.30
[C10] Studied with another student?	0.94	0.95
[C11] Attended a social or cultural event with another student?	0.97	0.95
[C12] Helped another student with a personal problem they were having?	0.95	0.95
[C13] Helped another student by tutoring or teaching them?	1.02	1.01
[C14] Talked about news or current events?	0.99	0.95
[C15] Asked for advice about a personal concern or issue?	0.74	0.73

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
During your final year in high school, how often did you do the following with fellow students...					
[C1] Shared your concerns about a class, test, or assignment?	-0.56	-0.15	0.19	0.53	1.03
[C2] Discussed something you learned in class?	-0.87	-0.36	0.09	0.52	1.13
[C3] Had a discussion with someone whose political opinions were very different than your own?	-0.59	-0.06	0.28	0.59	1.20
[C4] Had a discussion with someone with very different religious beliefs than your own?	-0.45	0.01	0.39	0.72	1.24
[C5] Talked about social issues such as peace, justice, or human rights?	-0.58	-0.08	0.28	0.58	1.12
[C6] Discussed views about multiculturalism or diversity?	-0.52	-0.12	0.25	0.64	1.21
[C7] Had discussions with students whose personal values, [religious beliefs, or political opinions]* differed from your own?	-0.82	-0.17	0.21	0.44	1.03
[C8] Received advice about an academic issue?	-0.28	0.10	0.37	0.59	1.15
[C9] Worked on a group project with a class mate?	-0.25	0.15	0.39	0.67	1.00
[C10] Studied with another student?	-0.67	-0.21	0.12	0.47	0.92
[C11] Attended a social or cultural event with another student?	-0.40	0.09	0.42	0.60	1.14
[C12] Helped another student with a personal problem they were having?	-0.69	-0.19	0.14	0.45	0.91
[C13] Helped another student by tutoring or teaching them?	-0.26	0.03	0.32	0.69	1.17

Item During your final year in high school, how often did you do the following with fellow students...	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C14] Talked about news or current events?	-0.56	-0.15	0.19	0.53	1.03
[C15] Asked for advice about a personal concern or issue?	-0.87	-0.36	0.09	0.52	1.13

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.21
2	-0.66
3	0.56
4	1.31

### ***Summary***

This scale functions well as a measure of Peer Interaction, High School. There is evidence that the scale measures one underlying latent trait and is comprised of items that each contribute meaningfully to the scale. Rasch reliability of 0.73 indicates that the scale reliably differentiates among people of higher and lower ability levels.

## 11. Faculty Interaction, High School [T0FIHS]

### *Items*

This scale consists of ten items that ask students about their academic interaction with faculty during the final year of high school. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic faculty interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Four items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T0	Cohort 2016, T0
During your final year in high school, how often did you do the following with teachers...		
[E1] Visited informally before or after class about course content?	✓	✓
[E2] Talked about ideas from a class outside of class time?	✓	✓
[E3] Asked (or emailed) a teacher for information about a class?	✓	✓
[E5] Discussed personal problems or concerns?	✓	✓
[E6] Discussed college and/or career plans and ambitions?	✓	✓
[E7] Socialized informally at a school event?	✓	✓
[E11] Discussed course topics, ideas, or concepts outside of class?	X	✓
[E12] Discussed your academic performance?	X	✓
[E13] Gotten to know each other as people outside the classroom?	X	✓
[E14] Discussed issues of interest or importance to me outside the classroom?	X	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.11 to 1.06.

Item	Item Difficulty (logits)
During your final year in high school, how often did you do the following with teachers...	
[E6] Discussed college and/or career plans and ambitions?	-1.11
[E12] Discussed your academic performance?	-0.46
[E1] Visited informally before or after class about course content?	-0.09
[E7] Socialized informally at a school event?	-0.06
[E13] Gotten to know each other as people outside the classroom?	-0.02
[E3] Asked (or emailed) a teacher for information about a class?	0.00
[E2] Talked about ideas from a class outside of class time?	0.16
[E14] Discussed issues of interest or importance to me outside the classroom?	0.20
[E11] Discussed course topics, ideas, or concepts outside of class?	0.33
[E5] Discussed personal problems or concerns?	1.06

### ***Data and Psychometric Diagnostics***

Of the 2,225 people who provided data for this scale, 21 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 12%
- Category 2 = 21%
- Category 3 = 33%
- Category 4 = 23%
- Category 5 = 11%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.92
- Rasch reliability: 0.82

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 51.5% of the variance in the data (of the 51.8% of variance explained that would be expected if data are a good fit to the model).

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

<b>Item</b>	<b>Outfit MNSQ</b>	<b>Infit MNSQ</b>
During your final year in high school, how often did you do the following with fellow students...		
[E1] Visited informally before or after class about course content?	0.98	0.97
[E2] Talked about ideas from a class outside of class time?	0.82	0.81
[E3] Asked (or emailed) a teacher for information about a class?	1.21	1.21
[E5] Discussed personal problems or concerns?	1.09	1.10
[E6] Discussed college and/or career plans and ambitions?	0.97	0.96
[E7] Socialized informally at a school event?	1.13	1.14
[E11] Discussed course topics, ideas, or concepts outside of class?	0.68	0.68
[E12] Discussed your academic performance?	1.11	1.06
[E13] Gotten to know each other as people outside the classroom?	1.18	1.20
[E14] Discussed issues of interest or importance to me outside the classroom?	0.71	0.72

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item During your final year in high school, how often did you do the following with fellow students...	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E1] Visited informally before or after class about course content?	-1.63	-0.78	-0.03	0.72	1.60
[E2] Talked about ideas from a class outside of class time?	-1.67	-0.65	0.11	0.86	1.99
[E3] Asked (or emailed) a teacher for information about a class?	-1.33	-0.55	-0.07	0.61	1.66
[E5] Discussed personal problems or concerns?	-1.09	-0.14	0.43	1.13	2.18
[E6] Discussed college and/or career plans and ambitions?	-2.41	-1.23	-0.60	0.24	1.08
[E7] Socialized informally at a school event?	-1.55	-0.63	-0.08	0.62	1.49
[E11] Discussed course topics, ideas, or concepts outside of class?	-1.70	-0.64	0.18	0.91	2.02
[E12] Discussed your academic performance?	-1.84	-0.82	-0.25	0.38	1.27
[E13] Gotten to know each other as people outside the classroom?	-1.55	-0.51	-0.01	0.46	1.39
[E14] Discussed issues of interest or importance to me outside the classroom?	-1.73	-0.67	0.08	0.85	1.88

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.79
2	-0.83
3	0.75
4	1.87

### *Summary*

This scale functions well as a measure of Faculty Interaction, High School. There is evidence that the scale measures one underlying latent trait and is comprised of items that each contribute meaningfully to the scale. Rasch reliability of 0.82 indicates that the scale reliably differentiates among people of higher and lower ability levels.

## 12. Faculty Encouragement, High School [T0FEHS]

### *Items*

This scale consists of three items that ask students about academic encouragement from faculty during their final year in high school. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of faculty encouragement.

The response options for each item in this scale are:

- Strongly agree
- Agree
- Disagree
- Strongly disagree

### Changes Over Time:

There were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.27 to 1.27.

Item	Item Difficulty (logits)
In the next section, please indicate the extent to which you agree or disagree with the following statements about your teachers in <i>high school</i> .	
[E10] My teachers believed in my potential to succeed academically.	-1.27
[E9] My teachers encouraged me to ask questions and participate in discussions.	0.01
[E8] My teachers provided me with feedback that helped me assess my progress in class.	1.27

### *Data and Psychometric Diagnostics*

Of the 2,225 people who provided data for this scale, 553 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 3%
- Category 2 = 4%
- Category 3 = 47%
- Category 4 = 46%

### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.75
- Rasch reliability: 0.59

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 51.1% of the variance in the data (of the 51.5% of variance explained that would be expected if data are a good fit to the model). There is one large contrast accounting for 25.8% of unexplained variance, which may indicate multidimensionality in this measure.

Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
In the next section, please indicate the extent to which you agree or disagree with the following statements about your teachers in <i>high school</i> .		
[E8] My teachers provided me with feedback that helped me assess my progress in class.	0.95	1
[E9] My teachers encouraged me to ask questions and participate in discussions.	0.91	0.91
[E10] My teachers believed in my potential to succeed academically.	1.09	1.13

Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
In the next section, please indicate the extent to which you agree or disagree with the following statements about your teachers in <i>high school</i> .				
[E8] My teachers provided me with feedback that helped me assess my progress in class.	-3.02	-0.06	3.04	6.11
[E9] My teachers encouraged me to ask questions and participate in discussions.	-3.91	-0.4	2.32	5.6
[E10] My teachers believed in my potential to succeed academically.	-4.32	-0.57	1.7	5.02

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.70
2	-1.54
3	4.25

**Summary**

This scale functions well as a measure of Faculty Encouragement. There is evidence that the scale reliably differentiates among people of high and low ability levels and is comprised of



items that each contribute meaningfully to the scale; however, a large percentage of variance accounted for by a secondary factor may indicate that the items do not measure one latent trait.

## Follow-Up Survey Scales

### 13. Self-Efficacy: Social Adjustment [T#SESA]

#### Items

This scale consists of eight items that ask students about their social experiences at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Self-Efficacy: Social Adjustment.

The response options for each item in this scale are:

Cannot do this at all        Absolutely can do this

#### Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:					
[L.1] Fit in socially at {INSTITUTION}.	✓	✓	✓	✓	✓
[L.2] Get involved in interesting activities.	✓	✓	✓	✓	✓
[L.3] Adjust well to life as a college student.	✓	✓	✓	✓	✓
[L.4] Develop good relationships with other students.	✓	✓	✓	✓	✓
[L.5] Feel at ease with others.	✓	✓	✓	✓	✓
[L.6] Meet people and make friends.	✓	✓	✓	✓	✓
[L.7] Make friends you can talk about your very personal problems with.	X	✓	✓	✓	✓
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}.	X	✓	✓	✓	✓

#### Item Difficulties and Hierarchy:

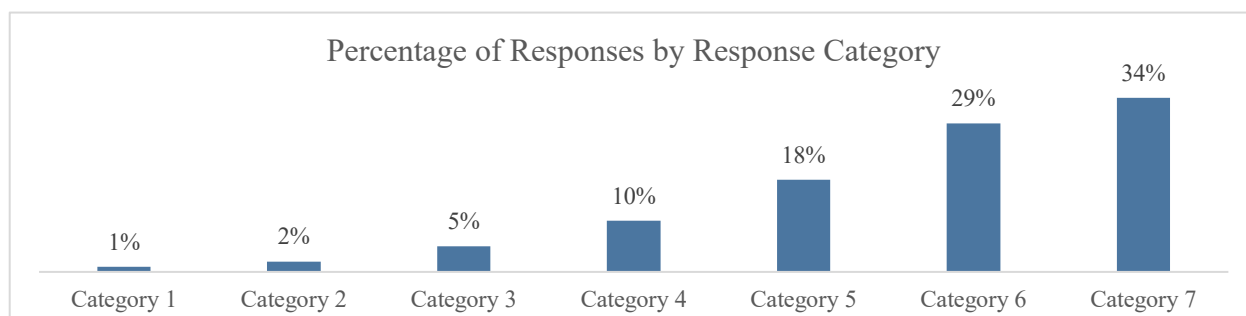
Item difficulties for this scale range from -0.34 to 0.75.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:	
[L.4] Develop good relationships with other students.	-0.34
[L.3] Adjust well to life as a college student.	-0.28
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}.	-0.19
[L.1] Fit in socially at {INSTITUTION}.	-0.16
[L.6] Meet people and make friends.	-0.06
[L.5] Feel at ease with others.	0.05
[L.2] Get involved in interesting activities.	0.23
[L.7] Make friends you can talk about your very personal problems with.	0.75

### ***Data and Psychometric Diagnostics***

Of the 4,344 people who provided data for this scale, 429 had maximum extreme scores (i.e., selected the most positive response option for all items), and 5 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.94
- Rasch reliability: 0.80

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 58.6% of the variance in the data (of the 58.1% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:		
[L.1] Fit in socially at {INSTITUTION}.	0.82	0.82
[L.2] Get involved in interesting activities.	1.13	1.07
[L.3] Adjust well to life as a college student.	1.15	1.10
[L.4] Develop good relationships with other students.	0.67	0.70
[L.5] Feel at ease with others.	0.75	0.69
[L.6] Meet people and make friends.	0.74	0.75
[L.7] Make friends you can talk about your very personal problems with.	1.28	1.26
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}.	1.56	1.50

Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[L.1] Fit in socially at {INSTITUTION}.	-2.12	-0.78	-0.21	0.25	0.90	1.71	3.56
[L.2] Get involved in interesting activities.	-2.13	-0.42	0.15	0.62	1.17	1.92	3.73
[L.3] Adjust well to life as a college student.	-2.08	-0.68	-0.15	0.41	0.91	1.62	3.49
[L.4] Develop good relationships with other students.	-2.74	-1.16	-0.35	0.08	0.72	1.56	3.52
[L.5] Feel at ease with others.	-2.96	-0.88	-0.28	0.33	1.02	1.88	3.93
[L.6] Meet people and make friends.	-2.36	-0.74	-0.15	0.29	0.96	1.72	3.62
[L.7] Make friends you can talk about your very personal problems with.	-0.72	0.10	0.48	1.01	1.48	2.09	4.02
[L.8] Be completely satisfied with your decision to attend {INSTITUTION}.	-1.35	-0.20	0.15	0.48	1.07	1.64	3.23

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.95
2	-1.00
3	-0.58
4	0.13
5	0.96
6	2.45

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

**Summary**

This scale functions well as a measure of Self-Efficacy: Social Adjustment. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 14. Self-Efficacy: Academic Adjustment [T#SEAA]

### Items

This scale consists of 14 items that ask students about their academic experiences at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Self-Efficacy: Academic Adjustment.

The response options for each item in this scale are:

Cannot do this at all ○ ○ ○ ○ ○ ○ Absolutely can do this

### Changes Over Time:

Three items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:					
[M.1] Finish my assignments on time.	✓	✓	✓	✓	✓
[M.2] Get myself to study even when there are other interesting things to do.	✓	✓	✓	✓	✓
[M.3] Take good notes during class.	✓	✓	✓	✓	✓
[M.4] Organize my schoolwork.	✓	✓	✓	✓	✓
[M.5] Remember information presented in class.	✓	✓	✓	✓	✓
[M.6] Remember information presented in textbooks.	✓	✓	✓	✓	✓
[M.7] Arrange a place to study without distractions.	✓	✓	✓	✓	✓
[M.8] Motivate myself to do schoolwork.	✓	✓	✓	✓	✓
[M.10] Understand what my professors expect of me academically.	✓	✓	✓	✓	✓
[M.11] Have effective study skills.	✓	✓	✓	✓	✓
[M.12] Meet the academic demands of college.	✓	✓	✓	✓	✓
[M.16] Work your hardest to do as well as you can.	X	✓	✓	✓	✓
[M.17] Be completely satisfied with your academic performance.	X	✓	✓	✓	✓
[M.18] Always keep up-to-date with your schoolwork.	X	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.62 to 0.80.

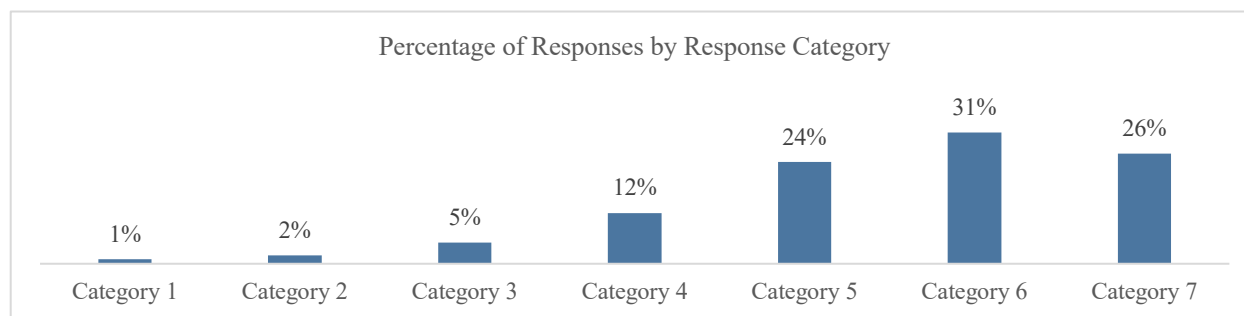
Item	Item Difficulty (logits)
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:	
[M.1] Finish my assignments on time.	-0.62
[M.16] Work your hardest to do as well as you can.	-0.51
[M.10] Understand what my professors expect of me academically.	-0.34
[M.4] Organize my schoolwork.	-0.27

Item	Item Difficulty (logits)
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:	
[M.12] Meet the academic demands of college.	-0.22
[M.18] Always keep up-to-date with your schoolwork.	-0.19
[M.7] Arrange a place to study without distractions.	-0.17
[M.3] Take good notes during class.	0.00
[M.5] Remember information presented in class.	0.20
[M.11] Have effective study skills.	0.25
[M.8] Motivate myself to do schoolwork.	0.26
[M.17] Be completely satisfied with your academic performance.	0.30
[M.2] Get myself to study even when there are other interesting things to do.	0.52
[M.6] Remember information presented in textbooks.	0.80

### ***Data and Psychometric Diagnostics***

Of the 4,344 people who provided data for this scale, 167 had maximum extreme scores (i.e., selected the most positive response option for all items), and 2 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### **Reliability Coefficients:**

- Cronbach's  $\alpha$  (inter-item consistency): 0.96
- Rasch reliability: 0.88

#### **Data Fit to Model and Dimensionality:**

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 55.0% of the variance in the data (of the 55.9% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:		
[M.1] Finish my assignments on time.	1.12	1.22
[M.2] Get myself to study even when there are other interesting things to do.	1.09	1.05
[M.3] Take good notes during class.	1.13	1.16
[M.4] Organize my schoolwork.	0.98	1.02
[M.5] Remember information presented in class.	0.88	0.88
[M.6] Remember information presented in textbooks.	1.36	1.29
[M.7] Arrange a place to study without distractions.	1.29	1.33
[M.8] Motivate myself to do schoolwork.	0.84	0.85
[M.10] Understand what my professors expect of me academically.	1.07	1.09
[M.11] Have effective study skills.	0.73	0.73
[M.12] Meet the academic demands of college.	0.80	0.83
[M.16] Work your hardest to do as well as you can.	0.95	1.00
[M.17] Be completely satisfied with your academic performance.	0.99	0.95
[M.18] Always keep up-to-date with your schoolwork.	0.84	0.84

Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:							
[M.1] Finish my assignments on time.	-1.84	-0.79	-0.11	0.28	0.91	1.54	2.92
[M.2] Get myself to study even when there are other interesting things to do.	-1.11	-0.16	0.38	0.84	1.46	2.17	3.85
[M.3] Take good notes during class.	-1.00	-0.17	0.08	0.52	1.19	1.84	3.45
[M.4] Organize my schoolwork.	-2.46	-0.49	0.01	0.38	0.95	1.76	3.26
[M.5] Remember information presented in class.	-2.17	-0.29	0.06	0.55	1.26	2.05	4.03
[M.6] Remember information presented in textbooks.	-1.03	0.43	0.56	0.99	1.60	2.36	4.37
[M.7] Arrange a place to study without distractions.	-1.25	-0.29	0.19	0.50	1.12	1.80	3.15
[M.8] Motivate myself to do schoolwork.	-1.33	-0.31	0.12	0.57	1.23	2.03	3.86
[M.10] Understand what my professors expect of me academically.	-1.09	-0.73	-0.01	0.36	0.94	1.73	3.25
[M.11] Have effective study skills.	-1.45	-0.49	-0.07	0.54	1.21	2.07	4.04
[M.12] Meet the academic demands of college.	-2.02	-0.60	-0.12	0.32	0.97	1.78	3.51
[M.16] Work your hardest to do as well as you can.	-2.32	-0.80	-0.28	0.23	0.86	1.64	3.15
[M.17] Be completely satisfied with your academic performance.	-1.01	-0.26	0.22	0.59	1.28	2.08	3.95
[M.18] Always keep up-to-date with your schoolwork.	-2.42	-0.64	-0.15	0.30	1.02	1.77	3.48

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options (i.e., respondents do not necessarily see each as distinct). Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.31
2	-1.17
3	-0.61
4	0.09
5	1.25
6	2.75

Differential Item Functioning:

For each of the 14 items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions well as a measure of Self-Efficacy: Academic Adjustment. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.



## 15. Mattering: Campus [T#MATCAMP]

### Items

This scale consists of eight items that ask students about supportive relationships they have at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Mattering: Campus.

The response options for each item in this scale are<sup>10</sup>:

Strongly Disagree                                Strongly Agree

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1 (4-point ROs)	Cohort 2015, T2 (7-point ROs)	Cohort 2015, T3 (7-point ROs)	Cohort 2016, T1 (7-point ROs)	Cohort 2016, T2 (7-point ROs)
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:					
[O.1] There are people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	✓	✓	✓	✓	✓
[O.3] There are people at {INSTITUTION} who are generally supportive of my individual needs.	✓	✓	✓	✓	✓
[O.4] There are people at {INSTITUTION} who seem happy about my accomplishments.	✓	✓	✓	✓	✓
[O.6] There are people at {INSTITUTION} who are concerned about my future.	✓	✓	✓	✓	✓
[O.7] There are people at {INSTITUTION} who are genuinely interested in me as a person.	✓	✓	✓	✓	✓
[O.8] Other students at {INSTITUTION} are happy for me when I do well on exams or projects.	✓	✓	✓	✓	✓
[O.15] People I value at {INSTITUTION} are disappointed when I don't accomplish all I should.	X	✓	✓	✓	✓
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	X	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.56 to 0.93.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[O.4] There are people at {INSTITUTION} who seem happy about my accomplishments.	-0.74
[O.3] There are people at {INSTITUTION} who are generally supportive of my individual needs.	-0.56
[O.7] There are people at {INSTITUTION} who are genuinely interested in me as a person.	-0.41

<sup>10</sup> Students in Cohort 2015 were presented with four response options in timepoint 1. To be scaled together, data from subsequent administrations were collapsed and scale diagnostics were examined with 4-point response options. See the Methods section for more information.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[O.8] Other students at {INSTITUTION} are happy for me when I do well on exams or projects.	-0.40
[O.6] There are people at {INSTITUTION} who are concerned about my future.	-0.18
[O.1] There are people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	0.61
[O.15] People I value at {INSTITUTION} are disappointed when I don't accomplish all I should.	0.75
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	0.93

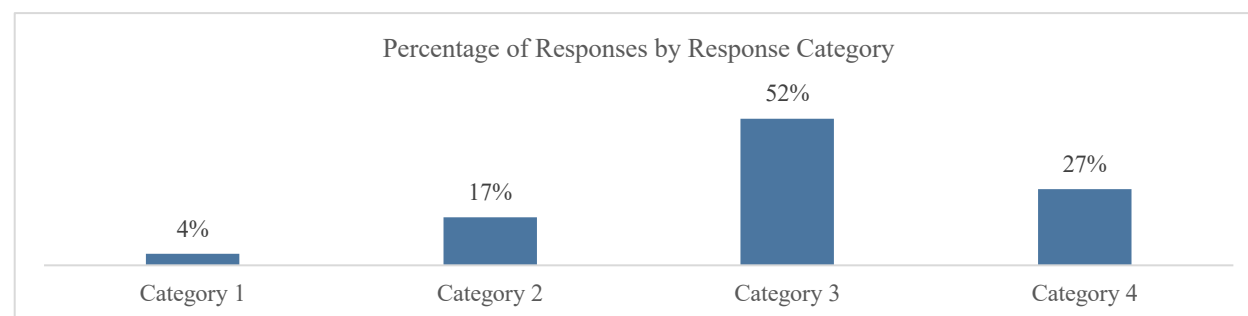
### *Data and Psychometric Diagnostics*

Of the 4,344 people who provided data for this scale, 329 had maximum extreme scores (i.e., selected the most positive response option for all items), and 28 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.85

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 56.3% of the variance in the data (of the 56.0% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[O.1] There are people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	1.01	0.99
[O.3] There are people at {INSTITUTION} who are generally supportive of my individual needs.	0.75	0.77
[O.4] There are people at {INSTITUTION} who seem happy about my accomplishments.	0.69	0.73
[O.6] There are people at {INSTITUTION} who are concerned about my future.	1.05	1.07
[O.7] There are people at {INSTITUTION} who are genuinely interested in me as a person.	0.79	0.83
[O.8] Other students at {INSTITUTION} are happy for me when I do well on exams or projects.	0.89	0.92
[O.15] People I value at {INSTITUTION} are disappointed when I don't accomplish all I should.	1.08	1.04
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	1.59	1.52

Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:				
[O.1] There are people at {INSTITUTION} who are sad for me when I fail in something I set out to do.	-2.14	-0.24	1.80	4.93
[O.3] There are people at {INSTITUTION} who are generally supportive of my individual needs.	-3.87	-1.10	1.25	4.36
[O.4] There are people at {INSTITUTION} who seem happy about my accomplishments.	-4.13	-1.18	1.13	4.26
[O.6] There are people at {INSTITUTION} who are concerned about my future.	-2.30	-0.65	1.33	4.50
[O.7] There are people at {INSTITUTION} who are genuinely interested in me as a person.	-3.33	-0.94	1.35	4.42
[O.8] Other students at {INSTITUTION} are happy for me when I do well on exams or projects.	-3.48	-0.82	1.36	4.43
[O.15] People I value at {INSTITUTION} are disappointed when I don't accomplish all I should.	-2.15	-0.16	1.85	5.26
[O.16] I feel pressure to do my best because people I value at {INSTITUTION} otherwise would be disappointed.	-1.35	0.14	1.90	4.87

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-3.02
2	-0.56
3	3.58

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions well as a measure of Mattering: Campus. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 16. Sense of Belonging: Campus [T#SOBCAMP]

### Items

This scale consists of eight items that ask students about their sense of belonging at their institution. Five of the eight items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and three of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of sense of belonging.

The response options for each item in this scale are<sup>11</sup>:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1 (4-point ROs)	Cohort 2015, T2 (7-point ROs)	Cohort 2015, T3 (7-point ROs)	Cohort 2016, T1 (7-point ROs)	Cohort 2016, T2 (7-point ROs)
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:					
[Q.1] I feel like an outsider.*	✓	✓	✓	✓	✓
[Q.2] I make friends easily.	✓	✓	✓	✓	✓
[Q.3] I feel like I belong.	✓	✓	✓	✓	✓
[Q.4] I feel awkward and out of place.*	✓	✓	✓	✓	✓
[Q.5] I feel lonely.*	✓	✓	✓	✓	✓
[Q.6] I believe other students like me.	✓	✓	✓	✓	✓
[Q.7] I see myself as an important part of the {INSTITUTION} community.	X	✓	✓	✓	✓
[Q.8] I feel I am a member of the {INSTITUTION} community.	X	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.53 to 0.46.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[Q.6] I believe other students like me.	-0.53
[Q.3] I feel like I belong.	-0.30
[Q.8] I feel I am a member of the {INSTITUTION} community.	-0.14
[Q.2] I make friends easily.	0.05
[Q.5] I feel lonely.*	0.08

<sup>11</sup> Students in Cohort 2015 were presented with four response options in timepoint 1. To be scaled together, data from subsequent administrations were collapsed, and scale diagnostics were examined with 4-point response options. See the Methods section for more information.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[Q.1] I feel like an outsider.*	0.12
[Q.4] I feel awkward and out of place.*	0.27
[Q.7] I see myself as an important part of the {INSTITUTION} community.	0.46

\*Negatively valenced item; reverse coded.

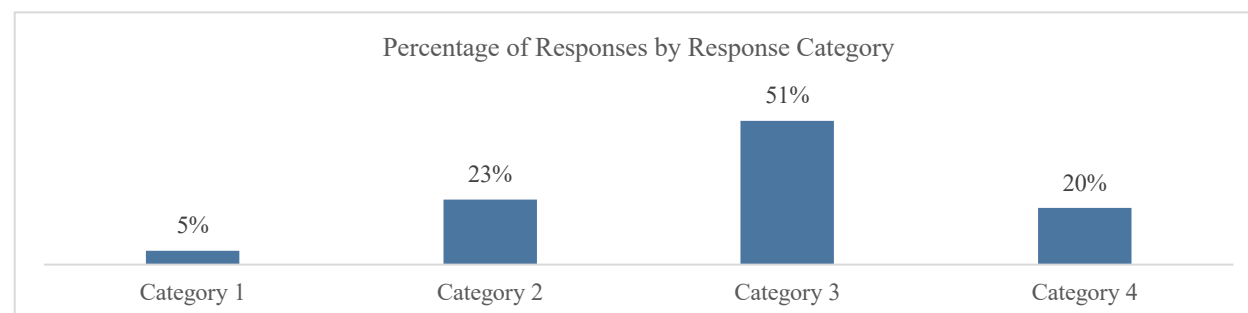
### *Data and Psychometric Diagnostics*

Of the 4,344 people who provided data for this scale, 196 had maximum extreme scores (i.e., selected the most positive response option for all items), and 16 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Across all eight items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.90
- Rasch reliability: 0.84

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 53.2% of the variance in the data (of the 52.2% of variance explained that would be expected if data are

a good fit to the model). The largest contrast accounts for 12.7% of unexplained variance, a slightly notable percentage but substantially lower than the variance accounted for by the measure. Not uncommon in scales with positively and negatively valenced items, the largest contrast is defined by the positively valenced and negatively valenced items in this scale. There are no other large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[Q.1] I feel like an outsider.*	0.98	0.96
[Q.2] I make friends easily.	1.12	1.13
[Q.3] I feel like I belong.	0.69	0.70
[Q.4] I feel awkward and out of place.*	0.90	0.88
[Q.5] I feel lonely.*	1.07	1.07
[Q.6] I believe other students like me.	0.83	0.82
[Q.7] I see myself as an important part of the {INSTITUTION} community.	1.10	1.09
[Q.8] I feel I am a member of the {INSTITUTION} community.	1.02	1.03

\*Negatively valenced item; reverse coded.

#### Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:				
[Q.1] I feel like an outsider.*	-1.56	-0.58	1.02	3.52
[Q.2] I make friends easily.	-1.97	-0.48	1.16	3.45
[Q.3] I feel like I belong.	-2.84	-0.95	0.98	3.63
[Q.4] I feel awkward and out of place.*	-1.70	-0.53	1.17	3.72
[Q.5] I feel lonely.*	-1.68	-0.43	1.05	3.29
[Q.6] I believe other students like me.	-2.62	-1.14	0.91	3.55
[Q.7] I see myself as an important part of the {INSTITUTION} community.	-1.94	-0.29	1.31	3.68
[Q.8] I feel I am a member of the {INSTITUTION} community.	-2.13	-0.71	0.99	3.46

\*Negatively valenced item; reverse coded.

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

	Positively Valenced Response Scale	Negatively Valenced Response Scale
Category	Threshold	Threshold
1	-2.84	-2.13
2	-0.40	-0.39
3	3.22	2.52

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions well as a measure of Sense of Belonging: Campus. There is evidence that the scale measures one underlying latent trait, despite a small contrast attributed to the mix of positively and negatively valenced items. Furthermore, there is evidence that this scale reliably differentiates among people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.



## 17. Mattering: Learning Community [T#MATLC]

### *Items*

This scale consists of eight items that ask students about supportive relationships they have at their learning community. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Mattering: Learning Community.

The response options for each item in this scale are:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale

### Item Difficulties and Hierarchy:

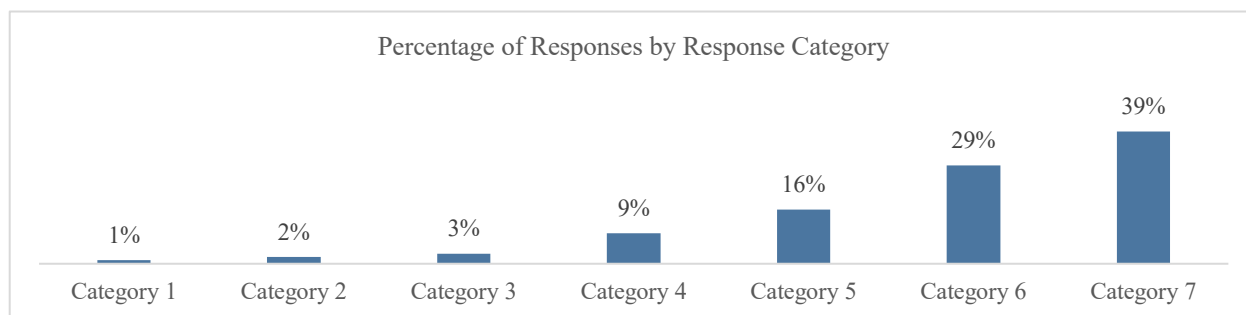
Item difficulties for this scale range from -0.54 to 0.60.

Item	Item Difficulty (logits)
Thinking about your experience as a student in the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[O.11] There are people at the {COMMUNITY} who seem happy about my accomplishments.	-0.54
[O.10] There are people at the {COMMUNITY} who are generally supportive of my individual needs.	-0.32
[O.13] There are people at the {COMMUNITY} who are genuinely interested in me as a person.	-0.23
[O.12] There are people at the {COMMUNITY} who are concerned about my future.	-0.20
[O.14] Other students at the {COMMUNITY} are happy for me when I do well on exams or projects.	-0.08
[O.9] There are people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	0.20
[O.17] People I value at the {COMMUNITY} are disappointed when I don't accomplish all I should.	0.57
[O.18] I feel pressure to do my best because people I value at the {COMMUNITY} otherwise would be disappointed.	0.60

### *Data and Psychometric Diagnostics*

Of the 1,254 people who provided data for this scale, 171 had maximum extreme scores (i.e., selected the most positive response option for all items), and 0 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all eight items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1, 2, and 3 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.90
- Rasch reliability: 0.77

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 54.4% of the variance in the data (of the 56.3% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for 12.7% of unexplained variance, a slightly notable percentage but substantially lower than the variance accounted for by the measure (see notes in the Summary section).

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student in the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[O.9] There are people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	0.89	0.83
[O.10] There are people at the {COMMUNITY} who are generally supportive of my individual needs.	0.71	0.73
[O.11] There are people at the {COMMUNITY} who seem happy about my accomplishments.	0.67	0.67
[O.12] There are people at the {COMMUNITY} who are concerned about my future.	1.24	1.23
[O.13] There are people at the {COMMUNITY} who are genuinely interested in me as a person.	0.90	0.92
[O.14] Other students at the {COMMUNITY} are happy for me when I do well on exams or projects.	0.99	0.93
[O.17] People I value at the {COMMUNITY} are disappointed when I don't accomplish all I should.	1.21	1.17
[O.18] I feel pressure to do my best because people I value at the {COMMUNITY} otherwise would be disappointed.	1.72	1.57

Ability Mean by Category:

With the exception of the values noted in red below, the data for this survey follow the expected pattern of average person ability measures by response categories. Deviation from the expected pattern may indicate that the response scale is not functioning as intended for those items.

Average person ability measures by response category are as follows:

Item Thinking about your experience as a student in the {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[O.9] There are people at the {COMMUNITY} who are sad for me when I fail in something I set out to do.	-0.07	-0.38	0.04	0.40	0.92	1.84	4.00
[O.10] There are people at the {COMMUNITY} who are generally supportive of my individual needs.	-1.94	-0.75	-0.21	0.06	0.71	1.47	3.51
[O.11] There are people at the {COMMUNITY} who seem happy about my accomplishments.	-2.08	-1.09	-0.66	0.02	0.50	1.38	3.39
[O.12] There are people at the {COMMUNITY} who are concerned about my future.	0.24	-0.32	0.05	0.44	0.60	1.45	3.48
[O.13] There are people at the {COMMUNITY} who are genuinely interested in me as a person.	-0.65	-0.45	-0.26	0.16	0.78	1.54	3.45
[O.14] Other students at the {COMMUNITY} are happy for me when I do well on exams or projects.	-1.48	-0.25	-0.05	0.35	0.92	1.60	3.58
[O.17] People I value at the {COMMUNITY} are disappointed when I don't accomplish all I should.	0.25	0.20	0.26	0.76	1.12	1.94	4.36
[O.18] I feel pressure to do my best because people I value at the {COMMUNITY} otherwise would be disappointed.	-0.07	-0.38	0.04	0.40	0.92	1.84	4.00

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.82
2	-1.06
3	-0.59
4	0.11
5	0.91
6	2.43

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions reasonably well as a measure of Mattering: Learning Community. There is evidence that the scale measures one underlying latent trait, but a contrast that accounts for more than 10% of unexplained variance indicated that further examination of multidimensionality was warranted. To this point we examined an alternative version of this measure with items O.10, O.11, O.13, and O.14 (based on diagnostics for the full scale) and did not find notable improvements with regard to dimensionality; therefore, we did not recommend changes. As is, there is evidence that this scale reliably differentiates among people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.

## 18. Validation [T#VALID]

### *Items*

This scale consists of 12 items that ask students about support, recognition, and encouragement that they received and perceived at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of validation.

The response options for each item in this scale are:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

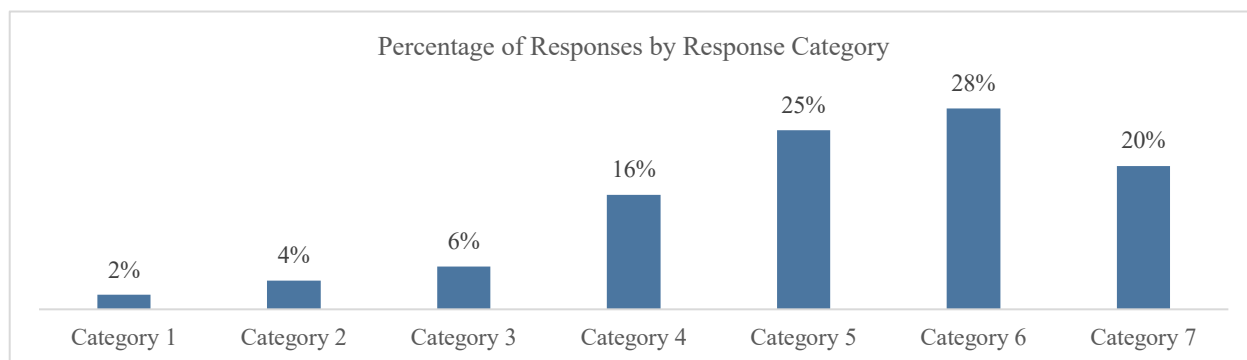
Item difficulties for this scale range from -0.46 to 0.39.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	-0.46
[OO.8] At least one faculty member has taken an interest in my development.	-0.24
[OO.7] Faculty believe in my potential to succeed academically.	-0.16
[OO.2] Instructors provided me with feedback that helped me judge my progress.	-0.14
[OO.9] At least one staff member has taken an interest in my development.	-0.12
[OO.1] I feel like my contributions were valued in class.	-0.06
[OO.11] Faculty empower me to learn here.	0.02
[OO.3] Instructors were able to determine my level of understanding of course material.	0.04
[OO.6] Instructors encouraged me to meet with them after or outside of class.	0.14
[OO.12] Staff encourage me to get involved in campus activities.	0.19
[OO.5] Instructors showed concern about my progress.	0.39
[OO.10] Staff recognize my achievements.	0.39

### *Data and Psychometric Diagnostics*

Of the 3,422 people who provided data for this scale, 150 had maximum extreme scores (i.e., selected the most positive response option for all items), and 2 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all 12 items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.88

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 55.7% of the variance in the data (of the 55.8% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[OO.1] I feel like my contributions were valued in class.	1.00	0.90
[OO.2] Instructors provided me with feedback that helped me judge my progress.	0.88	0.88
[OO.3] Instructors were able to determine my level of understanding of course material.	0.90	0.83
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	0.95	1.00
[OO.5] Instructors showed concern about my progress.	1.37	1.16
[OO.6] Instructors encouraged me to meet with them after or outside of class.	1.30	1.19
[OO.7] Faculty believe in my potential to succeed academically.	0.67	0.67
[OO.8] At least one faculty member has taken an interest in my development.	1.25	1.37
[OO.9] At least one staff member has taken an interest in my development.	1.22	1.28
[OO.10] Staff recognize my achievements.	0.90	0.86
[OO.11] Faculty empower me to learn here.	0.69	0.70
[OO.12] Staff encourage me to get involved in campus activities.	1.23	1.16

Ability Mean by Category:

With the exception of the value noted in red below, the data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[OO.1] I feel like my contributions were valued in class.	-1.40	-0.79	-0.33	0.12	0.72	1.50	3.45
[OO.2] Instructors provided me with feedback that helped me judge my progress.	-1.69	-0.72	-0.35	0.06	0.63	1.40	3.43
[OO.3] Instructors were able to determine my level of understanding of course material.	-1.71	-0.60	-0.25	0.10	0.79	1.55	3.75
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	-1.76	-0.91	-0.37	-0.11	0.48	1.22	2.92
[OO.5] Instructors showed concern about my progress.	-0.54	0.03	-0.09	0.37	0.83	1.66	3.91
[OO.6] Instructors encouraged me to meet with them after or outside of class.	-0.79	-0.38	-0.04	0.30	0.75	1.49	3.44
[OO.7] Faculty believe in my potential to succeed academically.	-1.67	-1.01	-0.47	0.01	0.60	1.39	3.40
[OO.8] At least one faculty member has taken an interest in my development.	-1.10	-0.54	-0.12	0.14	0.57	1.20	2.78
[OO.9] At least one staff member has taken an interest in my development.	-1.09	-0.48	-0.11	0.17	0.61	1.31	2.89
[OO.10] Staff recognize my achievements.	-1.06	-0.40	-0.01	0.36	0.92	1.68	3.84
[OO.11] Faculty empower me to learn here.	-1.57	-0.68	-0.29	0.10	0.69	1.53	3.59
[OO.12] Staff encourage me to get involved in campus activities.	-1.06	-0.34	-0.01	0.36	0.79	1.50	3.50

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are disordered, with no peak for category 3. In other words, there is no point in the ability distribution where response category 3 is the highest probability response). This finding, combined with the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.58
2	-0.90
3	-0.97
4	-0.03
5	0.92
6	2.50

Differential Item Functioning:

For each of the 12 items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions reasonably well as a measure of Validation. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.



## 19. Adapted Perceived Academic Control [T#APAC]

### Items

This scale consists of four items that ask students about their academic growth mindset. Two of the four items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and two of the four items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of adapted perceived academic control.

The response options for each item in this scale are:

Strongly Disagree, Disagree, Agree, Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
Thinking about the kind of student you are at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements...					
[S.1] I have a great deal of control over my academic performance.	✓	✓	✓	✓	✓
[S.2] The more effort I put into school, the better I do.	✓	✓	✓	✓	✓
[S.3] No matter what I do, I will not do well in school.	✓	✓	✓	✓	✓
[S.4] There is little I can do about my performance in school.	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.05 to 0.60.

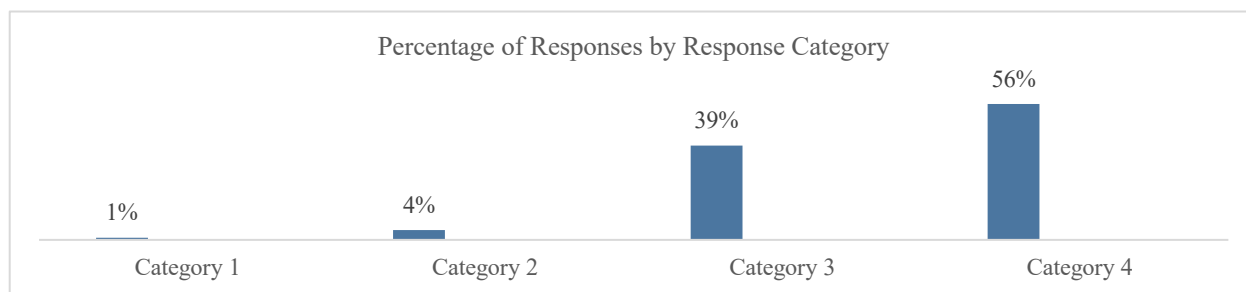
Item	Item Difficulty (logits)
Thinking about the kind of student you are at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements...	
[S.2] The more effort I put into school, the better I do.	-1.05
[S.1] I have a great deal of control over my academic performance.	0.11
[S.4] There is little I can do about my performance in school.*	0.34
[S.3] No matter what I do, I will not do well in school.*	0.60

\*Negatively valenced item; reverse coded.

### Data and Psychometric Diagnostics

Of the 4,344 people who provided data for this scale, 1,100 had maximum extreme scores (i.e., selected the most positive response option for all items), and 0 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



The minimal use of categories 1 and 2 may be an indication adding items of higher difficulty would improve this measure.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.72
- Rasch reliability: 0.53

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 42.8% of the variance in the data (of the 42.5% of variance explained that would be expected if data are a good fit to the model). The largest contrast accounts for a substantial percentage of unexplained variance (25.4%) but substantially lower than the variance accounted for by the measure, indicating some evidence of potential multidimensionality for future investigation. Not uncommon in scales with positively and negatively valenced items, the largest contrast is defined by the positively valenced and negatively valenced items in this scale. Other notable contrasts account for 16.6% and 15% of unexplained variance.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you are at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements...		
[S.1] I have a great deal of control over my academic performance.	0.93	0.97
[S.2] The more effort I put into school, the better I do.	1.19	1.18
[S.3] No matter what I do, I will not do well in school.	0.98	0.96
[S.4] There is little I can do about my performance in school.	0.87	0.90

#### Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item Thinking about the kind of student you are at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements...	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
[S.1] I have a great deal of control over my academic performance.	-0.90	-0.10	2.01	4.53
[S.2] The more effort I put into school, the better I do.	-1.09	-0.19	1.52	4.02
[S.3] No matter what I do, I will not do well in school.	-0.05	0.08	2.02	4.64
[S.4] There is little I can do about my performance in school.	-0.23	-0.07	1.86	4.58

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart for the positively valenced response scale. The minimal distance between thresholds in the negatively valenced response scale may indicate that response options were not well aligned to questions. Andrich thresholds for this scale are as follows:

	Positively Valenced Response Scale	Negatively Valenced Response Scale
Category	Threshold	Threshold
1	-2.12	-1.61
2	-1.12	-1.24
3	3.24	2.85

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item Thinking about the kind of student you are at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements...	Evidence of DIF (and item difficulty) Between...	Interpretation
[S.2] The more effort I put into school, the better I do.	White (-0.77) versus Black (-1.49), Asian (-1.67), and Hispanic (-1.41)	White students found this item more difficult to endorse compared with students of other races/ethnicities of comparable ability on other items.
[S.4] There is little I can do about my performance in school.	Asian (0.91) versus Hispanic (0.32) and White (0.25)	Asian students found this item more difficult to endorse compared with Hispanic and White students of comparable ability on other items.

### **Summary**

This scale functions acceptably as a measure of Adapted Perceived Academic Control, but there are areas of improvement as evidenced by diagnostic assessment of the scale's functioning. There is evidence that all items are well aligned to the overall construct and contribute

meaningful information; and that response options are relatively well suited to the items and construct, particularly for positively valenced items. However, there also is evidence that items are not well targeted to the population (i.e., are too easy): more than 25% of respondents selected the most positive response for each item, and overall respondents rarely selected the two least positive responses. This ceiling effect limits the measure's ability to distinguish respondents' ability on this measure (i.e., the scale has limited capacity to usefully distinguish between many levels of the latent trait). Finally, there also is evidence of possible multidimensionality that should be further explored.

## 20. Mentoring Support [T#MENTOR]

### *Items*

This scale consists of four items that ask students about their relationship with their mentor. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more interaction with mentors.

The response options for each item in this scale are:

Strongly Disagree, Disagree, Unsure, Agree, Strongly Agree, Not Applicable

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
Please rate how strongly you agree with each of the following statements.				
[AA.29] I actively participated in required one-on-one activities with my peer mentor.	✓	✓	✓	✓
[AA.30] I developed a close, personal relationship with my peer mentor.	✓	✓	✓	✓
[AA.31] I actively participated in required small-group activities led by my mentor.	✓	✓	✓	✓
[AA.32] I developed close, personal relationships with others in my small mentoring group.	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

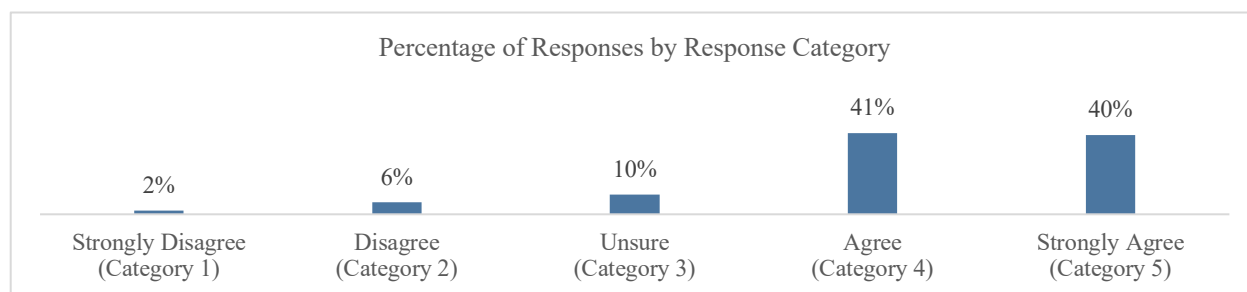
Item difficulties for this scale range from -1.30 to 1.03.

Item	Item Difficulty (logits)
Please rate how strongly you agree with each of the following statements.	
[AA.29] I actively participated in required one-on-one activities with my peer mentor.	-1.30
[AA.31] I actively participated in required small-group activities led by my mentor.	-0.11
[AA.30] I developed a close, personal relationship with my peer mentor.	0.38
[AA.32] I developed close, personal relationships with others in my small mentoring group.	1.03

### *Data and Psychometric Diagnostics*

Of the 1,242 people who provided data for this scale, 199 had maximum extreme scores (i.e., selected the most positive response option for all items), and 2 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.84
- Rasch reliability: 0.65

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 54.8% of the variance in the data (of the 55.1% of variance explained that would be expected if data are a good fit to the model). There is one notable contrast accounting for 18.2% of variance in the residuals separating item AA.32 from the other items. This contrast is not large enough to provide evidence for multidimensionality but should be monitored in future uses of the scale.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Please rate how strongly you agree with each of the following statements.		
[AA.29] I actively participated in required one-on-one activities with my peer mentor.	1.01	1.12
[AA.30] I developed a close, personal relationship with my peer mentor.	0.89	0.90
[AA.31] I actively participated in required small-group activities led by my mentor.	0.92	0.94
[AA.32] I developed close, personal relationships with others in my small mentoring group.	1.14	1.09

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Please rate how strongly you agree with each of the following statements.					
[AA.29] I actively participated in required one-on-one activities with my peer mentor.	-2.44	-1.51	.015	1.12	3.33

Item Please rate how strongly you agree with each of the following statements.	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[AA.30] I developed a close, personal relationship with my peer mentor.	-1.37	-0.08	0.80	1.82	4.33
[AA.31] I actively participated in required small-group activities led by my mentor.	-2.31	-0.24	0.61	1.57	4.09
[AA.32] I developed close, personal relationships with others in my small mentoring group.	-0.69	0.42	1.14	2.20	4.72

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.12
2	-0.48
3	-0.33
4	2.94

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender or race/ethnicity.

### **Summary**

This scale functions well as a measure of Mentoring Support. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options. However, Rasch reliability of 0.65 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 21. Engagement—First-Year Seminar Courses [T#ENGFYS]

### Items

This scale consists of 15 items<sup>12</sup> that ask students about their engagement in their first-year seminar. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more engagement.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item How often did you do the following in your {COURSE} at {INSTITUTION}?	Cohort 2015, T1	Cohort 2016, T1
[AA.2] Asked questions in class.	✓	✓
[AA.3] Participated in class discussions.	✓	✓
[AA.4] Worked on reading, writing, and/or problem-solving assignments during class.	✓	✓
[AA.5] Discussed complex topics with other students during class.	✓	✓
[AA.6] Provided feedback on other students' work during class.	✓	✓
[AA.7] Worked with other students on group projects during class.	✓	✓
[AA.8] Presented my work, or work done as part of a group, to the class.	✓	✓
[AA.9] Worked with classmates outside of class on class assignments, homework, or projects.	✓	✓
[AA.10] Worked to meet the high expectations of my instructors.	✓	✓
[AA.11] Developed friendships with classmates.	✓	✓
[AA.12] Took advantage of opportunities to interact with faculty outside of the classroom.	✓	✓
[AA.13] Used {MENTOR} to increase my academic performance in class. **	✓	✓
[AA.14] Used what I learned in other classes to help contribute to my success in this class.	✓	✓
[AA.15] Used what I learned in this class to help contribute to my success in other classes.\	✓	✓
[AA.16] Reflected on how these academic success strategies help me become a better student.	✓	✓

\*This item was only presented to UNO students.

<sup>12</sup> One item (AA.13) was only presented to students at the UNO campus. For this reason, scaling was tested with AA.13 excluded to examine whether the measurement properties of this scale improved when all items were consistent across campuses. The measurement properties were similar; therefore, the entire scale was kept intact, including item AA.13.



Item Difficulties and Hierarchy:

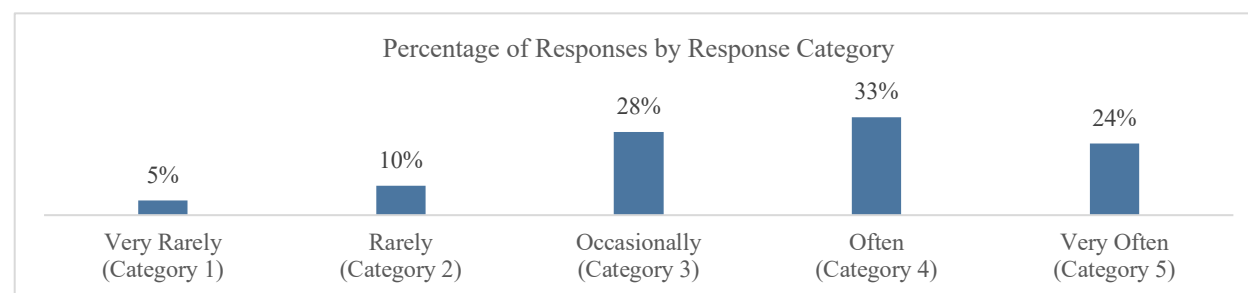
Item difficulties for this scale range from -0.74 to 1.23.

Item	Item Difficulty (logits)
How often did you do the following in your {COURSE} at {INSTITUTION}?	
[AA.10] Worked to meet the high expectations of my instructors.	-0.74
[AA.3] Participated in class discussions.	-0.49
[AA.11] Developed friendships with classmates.	-0.34
[AA.4] Worked on reading, writing, and/or problem-solving assignments during class.	-0.29
[AA.15] Used what I learned in this class to help contribute to my success in other classes.	-0.23
[AA.13] Used {MENTOR} to increase my academic performance in class. **	-0.19
[AA.16] Reflected on how these academic success strategies help me become a better student.	-0.13
[AA.5] Discussed complex topics with other students during class.	-0.12
[AA.14] Used what I learned in other classes to help contribute to my success in this class.	-0.05
[AA.7] Worked with other students on group projects during class.	0.08
[AA.8] Presented my work, or work done as part of a group, to the class.	0.12
[AA.12] Took advantage of opportunities to interact with faculty outside of the classroom.	0.18
[AA.6] Provided feedback on other students' work during class.	0.46
[AA.2] Asked questions in class.	0.49
[AA.9] Worked with classmates outside of class on class assignments, homework, or projects.	1.23

***Data and Psychometric Diagnostics***

Of the 910 people who provided data for this scale, 12 had maximum extreme scores (i.e., selected the most positive response option for all items), and 0 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.91
- Rasch reliability: 0.84

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 42.3% of the variance in the data (of the 42.2% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often did you do the following in your {COURSE} at {INSTITUTION}?		
[AA.2] Asked questions in class.	1.01	0.94
[AA.3] Participated in class discussions.	1.03	0.94
[AA.4] Worked on reading, writing, and/or problem-solving assignments during class.	0.86	0.87
[AA.5] Discussed complex topics with other students during class.	0.75	0.75
[AA.6] Provided feedback on other students' work during class.	0.82	0.82
[AA.7] Worked with other students on group projects during class.	1.19	1.20
[AA.8] Presented my work, or work done as part of a group, to the class.	1.00	0.98
[AA.9] Worked with classmates outside of class on class assignments, homework, or projects.	1.33	1.34
[AA.10] Worked to meet the high expectations of my instructors.	0.97	0.96
[AA.11] Developed friendships with classmates.	1.02	1.03
[AA.12] Took advantage of opportunities to interact with faculty outside of the classroom.	1.12	1.11
[AA.13] Used {MENTOR} to increase my academic performance in class. **	1.84	1.78
[AA.14] Used what I learned in other classes to help contribute to my success in this class.	0.75	0.75
[AA.15] Used what I learned in this class to help contribute to my success in other classes.	0.87	0.89
[AA.16] Reflected on how these academic success strategies help me become a better student.	0.81	0.80

Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often did you do the following in your {COURSE} at {INSTITUTION}?					
[AA.2] Asked questions in class.	-0.45	0.12	0.60	1.11	2.09
[AA.3] Participated in class discussions.	-1.85	-0.24	0.25	0.74	1.53
[AA.4] Worked on reading, writing, and/or problem-solving assignments during class.	-0.93	-0.11	0.20	0.80	1.75
[AA.5] Discussed complex topics with other students during class.	-0.98	-0.19	0.28	0.82	1.98

Item How often did you do the following in your {COURSE} at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[AA.6] Provided feedback on other students' work during class.	-0.40	0.05	0.47	1.15	2.30
[AA.7] Worked with other students on group projects during class.	-0.07	0.07	0.31	0.93	1.79
[AA.8] Presented my work, or work done as part of a group, to the class.	-0.38	0.04	0.36	0.93	1.94
[AA.9] Worked with classmates outside of class on class assignments, homework, or projects.	0.06	0.53	0.76	1.19	2.51
[AA.10] Worked to meet the high expectations of my instructors.	-0.99	-0.20	0.13	0.64	1.47
[AA.11] Developed friendships with classmates.	-0.55	-0.17	0.22	0.71	1.67
[AA.12] Took advantage of opportunities to interact with faculty outside of the classroom.	-0.28	0.06	0.42	0.95	1.84
[AA.13] Used {MENTOR} to increase my academic performance in class. **	0.06	0.25	0.51	0.95	1.37
[AA.14] Used what I learned in other classes to help contribute to my success in this class.	-0.81	-0.17	0.30	0.85	2.02
[AA.15] Used what I learned in this class to help contribute to my success in other classes.	-0.64	-0.18	0.20	0.77	1.82
[AA.16] Reflected on how these academic success strategies help me become a better student.	-0.85	-0.05	0.20	0.87	1.91

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.13
2	-0.98
3	0.45
4	1.66

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item How often did you do the following in your {COURSE} at {INSTITUTION}?	Evidence of DIF (and item difficulty) Between...	Interpretation
[AA.10] Worked to meet the high expectations of my instructors.	Asian (-0.32) versus Black (0.04), Hispanic (-0.01), and White (0.06)	Asian students found this item easier to endorse compared with students of other race/ethnicities with comparable ability on other items.

***Summary***

This scale functions well as a measure of Engagement—First-Year Seminar Courses. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

## 22. Career Decision Making Self-Efficacy [T#CDMSE]

### *Items*

This scale consists of 15 items that ask students about their self-efficacy in regard to career decision making. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of their self-efficacy regarding making career decisions.

The response options for each item in this scale are:

Cannot do this at all, Highly uncertain can do, Uncertain can do, Certain can do, Highly certain can do

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item Please rate how certain you are that you can do the following...	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
[NN.1] Find more information about your planned occupation.	✓	✓	✓	✓	✓
[NN.2] Figure out whether your planned occupation will be “in demand” in the future.	✓	✓	✓	✓	✓
[NN.3] Write a good resume.	✓	✓	✓	✓	✓
[NN.4] Decide what’s important to you in a job.	✓	✓	✓	✓	✓
[NN.5] Learn the typical salary of people in the job you plan to hold.	✓	✓	✓	✓	✓
[NN.6] Prioritize competing work, life, and other demands.	✓	✓	✓	✓	✓
[NN.7] Talk with someone in your planned occupation or field.	✓	✓	✓	✓	✓
[NN.8] Identify potential employers for the type of job you plan to hold.	✓	✓	✓	✓	✓
[NN.9] Define a way of life that works for you.	✓	✓	✓	✓	✓
[NN.10] Find more information about graduate or professional school.	✓	✓	✓	✓	✓
[NN.11] Successfully navigate the process of interviewing for a job.	✓	✓	✓	✓	✓
[NN.12] Make a five year plan to achieve your goals.	✓	✓	✓	✓	✓
[NN.13] Make an accurate assessment of your skills and abilities.	✓	✓	✓	✓	✓
[NN.14] Figure out how to complete your chosen major.	✓	✓	✓	✓	✓
[NN.15] Figure out what job suits you best.	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

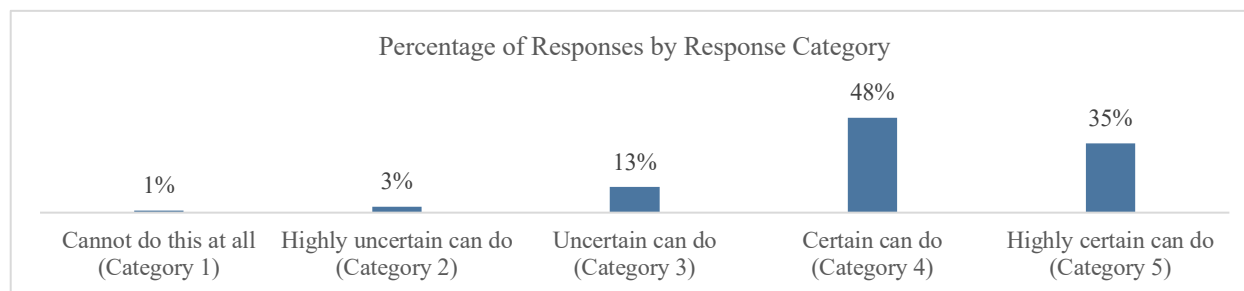
Item difficulties for this scale range from -0.69 to 0.41.

Item	Item Difficulty (logits)
Please rate how certain you are that you can do the following...	
[NN.14] Figure out how to complete your chosen major.	-0.69
[NN.5] Learn the typical salary of people in the job you plan to hold.	-0.31
[NN.4] Decide what's important to you in a job.	-0.23
[NN.7] Talk with someone in your planned occupation or field.	-0.18
[NN.10] Find more information about graduate or professional school.	-0.08
[NN.1] Find more information about your planned occupation.	-0.07
[NN.9] Define a way of life that works for you.	-0.02
[NN.11] Successfully navigate the process of interviewing for a job.	0.06
[NN.15] Figure out what job suits you best.	0.12
[NN.6] Prioritize competing work, life, and other demands.	0.12
[NN.13] Make an accurate assessment of your skills and abilities.	0.16
[NN.8] Identify potential employers for the type of job you plan to hold.	0.16
[NN.2] Figure out whether your planned occupation will be "in demand" in the future.	0.17
[NN.12] Make a five year plan to achieve your goals.	0.36
[NN.3] Write a good resume.	0.41

### ***Data and Psychometric Diagnostics***

Of the 4,346 people who provided data for this scale, 351 had maximum extreme scores (i.e., selected the most positive response option for all items), and 1 had a minimum extreme score (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### **Reliability Coefficients:**

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.86

### **Data Fit to Model and Dimensionality:**

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 44.0% of the variance in the data (of the 44.1% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Please rate how certain you are that you can do the following...		
[NN.1] Find more information about your planned occupation.	1.31	1.23
[NN.2] Figure out whether your planned occupation will be “in demand” in the future.	1.17	1.13
[NN.3] Write a good resume.	1.34	1.21
[NN.4] Decide what’s important to you in a job.	0.93	0.94
[NN.5] Learn the typical salary of people in the job you plan to hold.	0.98	0.98
[NN.6] Prioritize competing work, life, and other demands.	0.89	0.85
[NN.7] Talk with someone in your planned occupation or field.	1.01	1.02
[NN.8] Identify potential employers for the type of job you plan to hold.	0.97	0.96
[NN.9] Define a way of life that works for you.	0.83	0.84
[NN.10] Find more information about graduate or professional school.	0.95	0.95
[NN.11] Successfully navigate the process of interviewing for a job.	0.93	0.95
[NN.12] Make a five year plan to achieve your goals.	1.07	1.06
[NN.13] Make an accurate assessment of your skills and abilities.	0.85	0.86
[NN.14] Figure out how to complete your chosen major.	0.95	1.01
[NN.15] Figure out what job suits you best.	0.97	1.00

Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Please rate how certain you are that you can do the following...					
[NN.1] Find more information about your planned occupation.	0.09	0.44	0.68	1.78	4.04
[NN.2] Figure out whether your planned occupation will be “in demand” in the future.	-0.39	0.17	0.92	1.90	4.16
[NN.3] Write a good resume.	-0.24	0.19	1.16	2.06	4.25
[NN.4] Decide what’s important to you in a job.	-0.68	-0.35	0.56	1.76	4.06
[NN.5] Learn the typical salary of people in the job you plan to hold.	-0.91	-0.30	0.68	1.68	3.98
[NN.6] Prioritize competing work, life, and other demands.	-1.25	-0.23	0.85	1.87	4.42
[NN.7] Talk with someone in your planned occupation or field.	-0.81	-0.29	0.78	1.71	4.01
[NN.8] Identify potential employers for the type of job you plan to hold.	-0.79	-0.04	0.86	1.86	4.29
[NN.9] Define a way of life that works for you.	-1.38	-0.40	0.75	1.80	4.28
[NN.10] Find more information about graduate or professional school.	-1.09	-0.26	0.84	1.76	4.15
[NN.11] Successfully navigate the process of interviewing for a job.	-0.96	-0.21	0.92	1.80	4.30
[NN.12] Make a five year plan to achieve your goals.	-0.51	0.11	1.07	1.92	4.41

Item Please rate how certain you are that you can do the following...	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[NN.13] Make an accurate assessment of your skills and abilities.	-1.10	-0.18	0.89	1.85	4.49
[NN.14] Figure out how to complete your chosen major.	-1.64	-0.57	0.53	1.57	3.70
[NN.15] Figure out what job suits you best.	-0.88	-0.09	0.90	1.83	4.24

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.88
2	-1.25
3	0.05
4	3.08

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item Please rate how certain you are that you can do the following...	Evidence of DIF (and item difficulty) Between...	Interpretation
[NN.14] Figure out how to complete your chosen major.	White (-0.87) versus Black (-0.28)	Black students found this item more difficult to endorse compared with White students with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Career Decision Making Self-Efficacy. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.



## 23. Academic Peer Interaction, Non Learning Community [T#APINLC]

### *Items*

This scale consists of six items that ask students about their academic interaction with peers at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?					
[C.29] Shared your concerns about a class, test, or assignment?	✓	✓	✓	✓	✓
[C.30] Discussed something you learned in class?	✓	✓	✓	✓	✓
[C.34] Received advice about an academic issue?	✓	✓	✓	✓	✓
[C.35] Worked on a group project with a class mate?	X	✓	✓	✓	✓
[C.36] Studied with another student?	X	✓	✓	✓	✓
[C.39] Helped a classmate by tutoring or teaching them?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

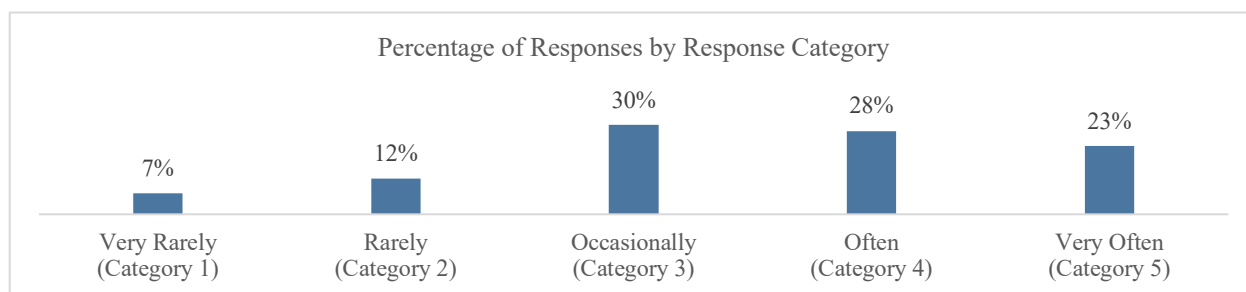
Item difficulties for this scale range from -0.58 to 1.07.

Item	Item Difficulty (logits)
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?	
[C.29] Shared your concerns about a class, test, or assignment?	-0.58
[C.30] Discussed something you learned in class?	-0.52
[C.35] Worked on a group project with a class mate?	-0.16
[C.34] Received advice about an academic issue?	0.09
[C.36] Studied with another student?	0.10
[C.39] Helped a classmate by tutoring or teaching them?	1.07

### *Data and Psychometric Diagnostics*

Of the 2,627 people who provided data for this scale, 88 had maximum extreme scores (i.e., selected the most positive response option for all items), and 5 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.87
- Rasch reliability: 0.73

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 51% of the variance in the data (of the 50.6% of variance explained that would be expected if data are a good fit to the model). There is one small contrast accounting for 14.5% of unexplained variance, which can be an area for future monitoring of this scale.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?		
[C.29] Shared your concerns about a class, test, or assignment?	1.09	1.08
[C.30] Discussed something you learned in class?	0.78	0.77
[C.35] Worked on a group project with a class mate?	0.80	0.77
[C.34] Received advice about an academic issue?	1.06	1.07
[C.36] Studied with another student?	1.11	1.13
[C.39] Helped a classmate by tutoring or teaching them?	1.05	1.04

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?					
[C.29] Shared your concerns about a class, test, or assignment?	-1.33	-0.57	0.06	0.70	1.76
[C.30] Discussed something you learned in class?	-1.79	-0.65	-0.04	0.74	2.01
[C.34] Received advice about an academic issue?	-1.24	-0.40	0.28	1.04	2.49
[C.35] Worked on a group project with a class mate?	-1.08	-0.27	0.21	0.83	2.05

Item How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C.36] Studied with another student?	-0.94	-0.15	0.34	0.90	2.10
[C.39] Helped a classmate by tutoring or teaching them?	-0.54	0.21	0.79	1.39	2.95

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.37
2	-0.98
3	0.67
4	1.69

### Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

### ***Summary***

This scale functions well as a measure of Academic Peer Interaction, Non learning Community. There is evidence that the scale measures one underlying latent trait and is comprised of items that each contribute meaningfully to the scale. Rasch reliability of 0.73 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 24. Academic Peer Interaction, Learning Community About Non Learning Community [T#APILCNLC]

### *Items*

This scale consists of six items that ask learning community students about their academic interaction with peers outside of their learning community at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?					
[C.1] Shared your concerns about a class, test, or assignment?	✓	✓	✓	✓	✓
[C.2] Discussed something you learned in class?	✓	✓	✓	✓	✓
[C.8] Received advice about an academic issue?	✓	✓	✓	✓	✓
[C.9] Worked on a group project with a class mate?	X	✓	✓	✓	✓
[C.10] Studied with another student?	X	✓	✓	✓	✓
[C.13] Helped a classmate by tutoring or teaching them?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.49 to 0.80.

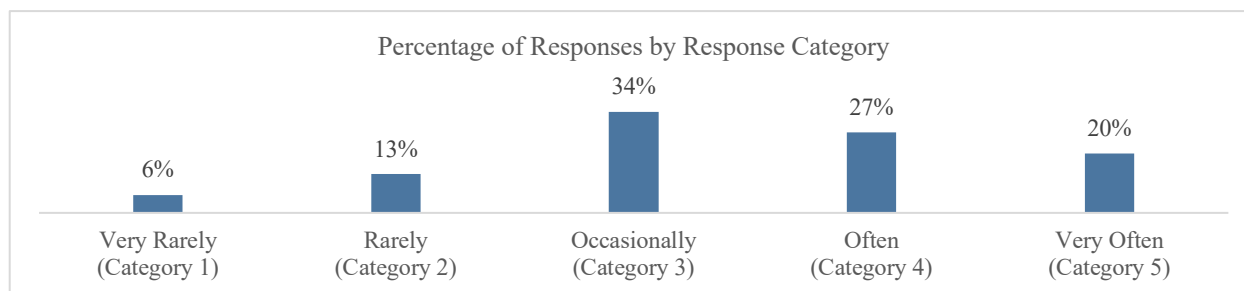
Item	Item Difficulty (logits)
How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?	
[C.2] Discussed something you learned in class?	-0.49
[C.1] Shared your concerns about a class, test, or assignment?	-0.28
[C.9] Worked on a group project with a class mate?	-0.10
[C.8] Received advice about an academic issue?	-0.02
[C.10] Studied with another student?	0.09
[C.13] Helped a classmate by tutoring or teaching them?	0.80

### *Data and Psychometric Diagnostics*

Of the 1,719 people who provided data for this scale, 79 had maximum extreme scores (i.e., selected the most positive response option for all items), and 9 had minimum extreme scores

(i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.87
- Rasch reliability: 0.75

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 49.6% of the variance in the data (of the 48.8% of variance explained that would be expected if data are a good fit to the model). There is one small contrast accounting for 14.4% of unexplained variance, which can be an area for future monitoring of this scale.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a non-{COMMUNITY} student during your {year} year at {INSTITUTION}?		
[C.1] Shared your concerns about a class, test, or assignment?	1.02	0.99
[C.2] Discussed something you learned in class?	0.76	0.74
[C.8] Received advice about an academic issue?	0.87	0.85
[C.9] Worked on a group project with a class mate?	1.03	1.01
[C.10] Studied with another student?	1.10	1.11
[C.13] Helped a classmate by tutoring or teaching them?	1.17	1.14

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C.1] Shared your concerns about a class, test, or assignment?	-1.51	-0.47	0.00	0.84	2.32
[C.2] Discussed something you learned in class?	-2.02	-0.81	-0.06	0.80	2.46
[C.8] Received advice about an academic issue?	-1.51	-0.49	0.19	1.05	2.76
[C.9] Worked on a group project with a class mate?	-1.25	-0.52	0.14	0.93	2.44
[C.10] Studied with another student?	-1.00	-0.36	0.15	0.91	2.50
[C.13] Helped a classmate by tutoring or teaching them?	-0.75	-0.02	0.59	1.40	3.09

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) and is at least one logit apart in distance. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.99
2	-0.90
3	0.70
4	2.18

### Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

### ***Summary***

This scale functions well as a measure of Academic Peer Interaction, Learning Community About Non learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. Rasch reliability of 0.75 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty

## 25. Academic Peer Interaction, Learning Community [T#APILC]

### Items

This scale consists of six items that ask learning community students about their academic interaction with peers in their learning community at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
[C.16] Shared your concerns about a class, test, or assignment?	✓	✓	✓	✓	✓
[C.17] Discussed something you learned in class?	✓	✓	✓	✓	✓
[C.21] Received advice about an academic issue?	✓	✓	✓	✓	✓
[C.22] Worked on a group project with a class mate?	X	✓	✓	✓	✓
[C.23] Studied with another student?	X	✓	✓	✓	✓
[C.26] Helped a classmate by tutoring or teaching them?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.43 to 0.84.

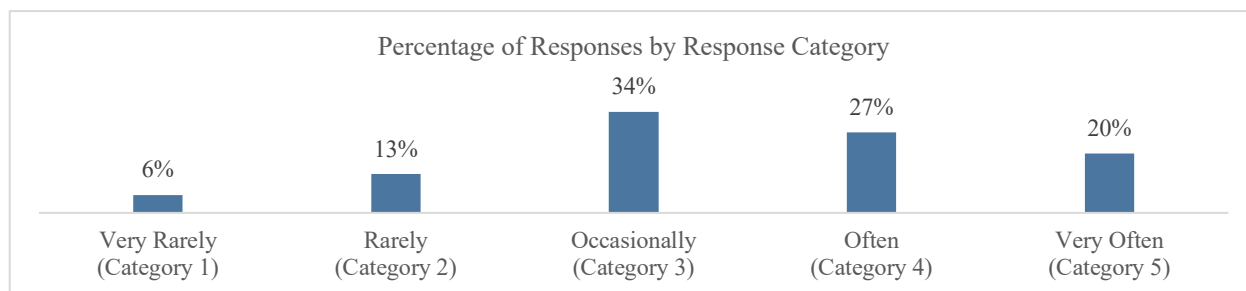
Item How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	Item Difficulty (logits)
[C.17] Discussed something you learned in class?	-0.43
[C.16] Shared your concerns about a class, test, or assignment?	-0.40
[C.21] Received advice about an academic issue?	-0.10
[C.22] Worked on a group project with a class mate?	0.04
[C.23] Studied with another student?	0.06
[C.26] Helped a classmate by tutoring or teaching them?	0.84

### Data and Psychometric Diagnostics

Of the 1,719 people who provided data for this scale, 119 had maximum extreme scores (i.e., selected the most positive response option for all items), and 20 had minimum extreme scores

(i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.90
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 55.9% of the variance in the data (of the 54.8% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?		
[C.16] Shared your concerns about a class, test, or assignment?	0.95	0.92
[C.17] Discussed something you learned in class?	0.76	0.77
[C.21] Received advice about an academic issue?	0.86	0.86
[C.22] Worked on a group project with a class mate?	1.11	1.10
[C.23] Studied with another student?	1.02	0.99
[C.26] Helped a classmate by tutoring or teaching them?	1.16	1.17

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?					
[C.16] Shared your concerns about a class, test, or assignment?	-2.43	-1.04	-0.05	0.97	2.74



Item How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C.17] Discussed something you learned in class?	-2.92	-1.04	-0.07	1.06	2.89
[C.21] Received advice about an academic issue?	-2.16	-0.74	0.18	1.30	3.12
[C.22] Worked on a group project with a class mate?	-1.78	-0.53	0.12	1.22	3.04
[C.23] Studied with another student?	-1.83	-0.61	0.16	1.22	3.05
[C.26] Helped a classmate by tutoring or teaching them?	-1.18	-0.14	0.69	1.65	3.59

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.90
2	-1.12
3	0.91
4	2.11

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	Evidence of DIF (and item difficulty) Between...	Interpretation
[C.22] Worked on a group project with a classmate?	Black (0.55) versus Asian (0.04), Hispanic (0.04), and White (0.04)	Black students found this item more difficult to endorse compared with students of other races/ethnicities of comparable ability on other items.

### **Summary**

This scale functions well as a measure of Academic Peer Interaction, Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options.

## 26. Social Peer Interaction, Non Learning Community [T#SPINLC]

### Items

This scale consists of seven items that ask students about their social interaction with peers at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of social peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Three items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?					
[C.31] Talked about social issues, such as peace, justice, or human rights?	✓	✓	✓	✓	✓
[C.32] Discussed views about multiculturalism or diversity?	X	✓	✓	✓	✓
[C.33] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	✓	✓	✓	✓	✓
[C.37] Attended a social or cultural event with another student?	X	✓	✓	✓	✓
[C.38] Helped another student with a personal problem they were having?	X	✓	✓	✓	✓
[C.40] Talked about current events or news?	✓	✓	✓	✓	✓
[C.41] Asked for advice about a personal concern or issue?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.32 to 0.42.

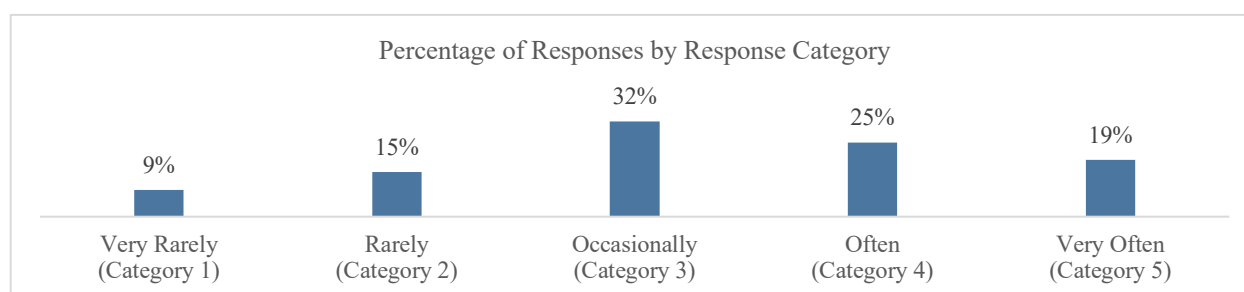
Item	Item Difficulty (logits)
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?	
[C.40] Talked about current events or news?	-0.32
[C.31] Talked about social issues, such as peace, justice, or human rights?	-0.20
[C.38] Helped another student with a personal problem they were having?	-0.10
[C.33] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	-0.05
[C.32] Discussed views about multiculturalism or diversity?	-0.01
[C.41] Asked for advice about a personal concern or issue?	0.26

Item	Item Difficulty (logits)
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?	
[C.37] Attended a social or cultural event with another student?	0.42

### ***Data and Psychometric Diagnostics***

Of the 2,627 people who provided data for this scale, 90 had maximum extreme scores (i.e., selected the most positive response option for all items), and 22 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.92
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 49.6% of the variance in the data (of the 48.5% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?		
[C.31] Talked about social issues, such as peace, justice, or human rights?	0.83	0.82
[C.32] Discussed views about multiculturalism or diversity?	0.87	0.88
[C.33] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	0.99	0.95
[C.37] Attended a social or cultural event with another student?	1.28	1.26
[C.38] Helped another student with a personal problem they were having?	0.90	0.90
[C.40] Talked about current events or news?	0.87	0.83

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?		
[C.41] Asked for advice about a personal concern or issue?	0.99	0.99

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following with a fellow student during your {year} year at {INSTITUTION}?					
[C.31] Talked about social issues, such as peace, justice, or human rights?	-1.82	-0.67	0.03	0.72	2.04
[C.32] Discussed views about multiculturalism or diversity?	-1.69	-0.55	0.13	0.75	2.11
[C.33] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	-1.71	-0.56	0.14	0.77	2.09
[C.37] Attended a social or cultural event with another student?	-1.09	-0.16	0.31	0.82	2.25
[C.38] Helped another student with a personal problem they were having?	-1.79	-0.63	0.09	0.77	2.03
[C.40] Talked about current events or news?	-2.01	-0.77	-0.03	0.67	2.02
[C.41] Asked for advice about a personal concern or issue?	-1.43	-0.33	0.17	0.87	2.21

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.39
2	-0.83
3	0.57
4	1.65

### Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

### **Summary**

This scale functions well as a measure of Social Peer Interaction, Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

## 27. Social Peer Interaction, Learning Community About Non Learning Community [T#SPILCNLC]

### Items

This scale consists of seven items that ask students about their social interaction with peers outside of the learning community at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of social peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Three items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?					
[C.5] Talked about social issues, such as peace, justice, or human rights?	✓	✓	✓	✓	✓
[C.6] Discussed views about multiculturalism or diversity?	X	✓	✓	✓	✓
[C.7] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	✓	✓	✓	✓	✓
[C.11] Attended a social or cultural event with another student?	X	✓	✓	✓	✓
[C.12] Helped another student with a personal problem they were having?	X	✓	✓	✓	✓
[C.14] Talked about current events or news?	✓	✓	✓	✓	✓
[C.15] Asked for advice about a personal concern or issue?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

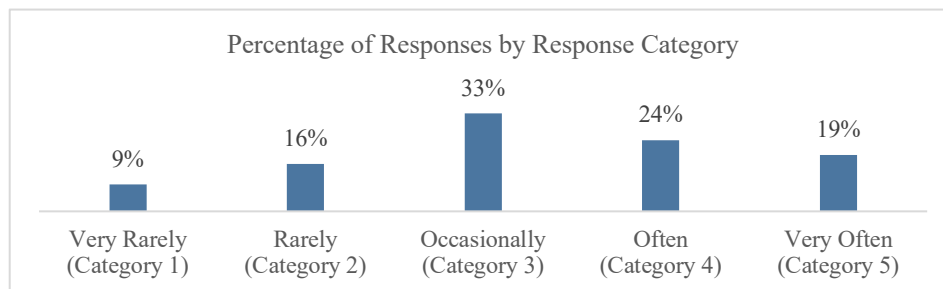
Item difficulties for this scale range from -0.15 to 0.30.

Item	Item Difficulty (logits)
How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?	
[C.12] Helped another student with a personal problem they were having?	-0.15
[C.14] Talked about current events or news?	-0.15
[C.5] Talked about social issues, such as peace, justice, or human rights?	-0.12
[C.7] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	-0.09
[C.6] Discussed views about multiculturalism or diversity?	-0.06
[C.15] Asked for advice about a personal concern or issue?	0.27
[C.11] Attended a social or cultural event with another student?	0.30

### ***Data and Psychometric Diagnostics***

Of the 1,719 people who provided data for this scale, 92 had maximum extreme scores (i.e., selected the most positive response option for all items), and 18 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.79

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 51% of the variance in the data (of the 50.3% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a non- <b>{COMMUNITY}</b> student during your <b>{year}</b> year at <b>{INSTITUTION}</b> ?		
[C.5] Talked about social issues, such as peace, justice, or human rights?	0.89	0.89
[C.6] Discussed views about multiculturalism or diversity?	0.90	0.90
[C.7] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	0.94	0.93
[C.11] Attended a social or cultural event with another student?	1.30	1.28
[C.12] Helped another student with a personal problem they were having?	0.94	0.94
[C.14] Talked about current events or news?	0.93	0.90
[C.15] Asked for advice about a personal concern or issue?	0.97	0.97

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a non- {COMMUNITY} student during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C.5] Talked about social issues, such as peace, justice, or human rights?	-1.79	-0.72	0.12	0.95	2.58
[C.6] Discussed views about multiculturalism or diversity?	-1.85	-0.73	0.06	0.85	2.53
[C.7] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	-2.02	-0.71	0.10	0.93	2.54
[C.11] Attended a social or cultural event with another student?	-1.40	-0.36	0.31	0.96	2.54
[C.12] Helped another student with a personal problem they were having?	-2.08	-0.65	0.01	0.86	2.40
[C.14] Talked about current events or news?	-1.89	-0.77	0.01	0.91	2.54
[C.15] Asked for advice about a personal concern or issue?	-1.56	-0.56	0.19	1.02	2.73

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.70
2	-0.94
3	0.81
4	1.82

### Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

### ***Summary***

This scale functions well as a measure of Social Peer Interaction, Learning Community About Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

## 28. Social Peer Interaction, Learning Community [T#SPILC]

### Items

This scale consists of seven items that ask students about their social interaction with peers in their learning community at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of social peer interactions.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Three items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?					
[C.18] Talked about social issues, such as peace, justice, or human rights?	✓	✓	✓	✓	✓
[C.19] Discussed views about multiculturalism or diversity?	X	✓	✓	✓	✓
[C.20] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	✓	✓	✓	✓	✓
[C.24] Attended a social or cultural event with another student?	X	✓	✓	✓	✓
[C.25] Helped another student with a personal problem they were having?	X	✓	✓	✓	✓
[C.27] Talked about current events or news?	✓	✓	✓	✓	✓
[C.28] Asked for advice about a personal concern or issue?	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.11 to 0.36.

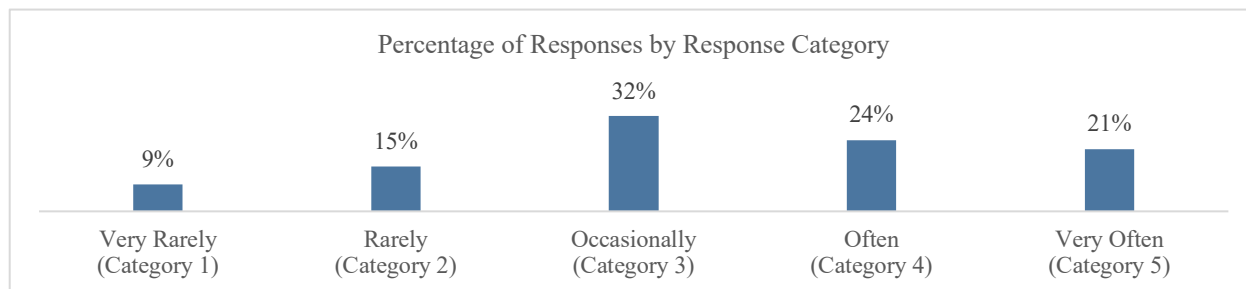
Item	Item Difficulty (logits)
How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	
[C.25] Helped another student with a personal problem they were having?	-0.11
[C.18] Talked about social issues, such as peace, justice, or human rights?	-0.09
[C.24] Attended a social or cultural event with another student?	-0.09
[C.27] Talked about current events or news?	-0.08
[C.19] Discussed views about multiculturalism or diversity?	-0.04
[C.20] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	0.05
[C.28] Asked for advice about a personal concern or issue?	0.36



### ***Data and Psychometric Diagnostics***

Of the 1,719 people who provided data for this scale, 134 had maximum extreme scores (i.e., selected the most positive response option for all items), and 27 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.94
- Rasch reliability: 0.81

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 52.3% of the variance in the data (of the 52.2% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?		
[C.18] Talked about social issues, such as peace, justice, or human rights?	0.84	0.85
[C.19] Discussed views about multiculturalism or diversity?	0.90	0.91
[C.20] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	1.08	1.07
[C.24] Attended a social or cultural event with another student?	1.23	1.18
[C.25] Helped another student with a personal problem they were having?	0.96	0.97
[C.27] Talked about current events or news?	0.90	0.88
[C.28] Asked for advice about a personal concern or issue?	1.08	1.09

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a fellow {COMMUNITY} student during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[C.18] Talked about social issues, such as peace, justice, or human rights?	-2.06	-0.80	0.10	1.10	3.11
[C.19] Discussed views about multiculturalism or diversity?	-2.19	-0.73	0.03	0.98	2.88
[C.20] Had discussions with classmates whose personal values, religious beliefs, or political opinions differed from your own?	-2.00	-0.64	0.09	1.05	2.99
[C.24] Attended a social or cultural event with another student?	-2.06	-0.72	0.08	0.96	2.74
[C.25] Helped another student with a personal problem they were having?	-2.19	-0.88	0.05	0.99	2.81
[C.27] Talked about current events or news?	-2.25	-0.88	0.05	1.11	3.02
[C.28] Asked for advice about a personal concern or issue?	-1.67	-0.49	0.21	1.20	3.08

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.75
2	-1.07
3	0.87
4	1.95

### Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

### ***Summary***

This scale functions well as a measure of Social Peer Interaction, Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

## 29. Engagement—Shared Academic Courses [T#ENGSAC]

### Items

This scale consists of 11 items that ask students about their engagement in their community-required academic courses. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more engagement.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
[AA.17] Asked questions in class.	✓	✓	✓	✓	✓
[AA.18] Participated in class discussions.	✓	✓	✓	✓	✓
[AA.19] Worked on reading, writing, and/or problem-solving assignments during class.	✓	✓	✓	✓	✓
[AA.20] Discussed complex topics with other students during class.	✓	✓	✓	✓	✓
[AA.21] Provided feedback on other students' work during class.	✓	✓	✓	✓	✓
[AA.22] Worked with other students on group projects during class.	✓	✓	✓	✓	✓
[AA.23] Presented my work, or work done as part of a group, to the class.	✓	✓	✓	✓	✓
[AA.24] Worked with classmates outside of class on class assignments, homework, or projects.	✓	✓	✓	✓	✓
[AA.25] Worked to meet the high expectations of my instructors.	✓	✓	✓	✓	✓
[AA.26] Developed friendships with classmates.	✓	✓	✓	✓	✓
[AA.27] Took advantage of opportunities to interact with faculty outside of the classroom.	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.51 to 0.28.

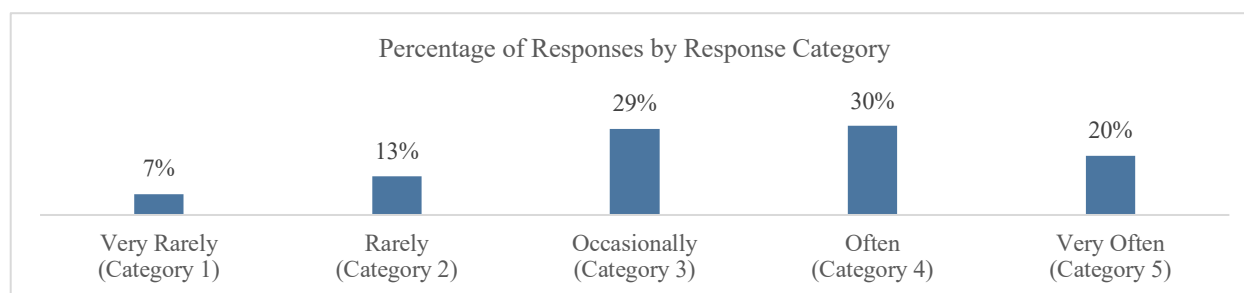
Item How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?	Item Difficulty (logits)
[AA.25] Worked to meet the high expectations of my instructors.	-0.51
[AA.19] Worked on reading, writing, and/or problem-solving assignments during class.	-0.18
[AA.18] Participated in class discussions.	-0.13

Item	Item Difficulty (logits)
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?	
[AA.26] Developed friendships with classmates.	-0.12
[AA.20] Discussed complex topics with other students during class.	-0.07
[AA.23] Presented my work, or work done as part of a group, to the class.	0.06
[AA.22] Worked with other students on group projects during class.	0.08
[AA.17] Asked questions in class.	0.12
[AA.21] Provided feedback on other students' work during class.	0.23
[AA.24] Worked with classmates outside of class on class assignments, homework, or projects.	0.25
[AA.27] Took advantage of opportunities to interact with faculty outside of the classroom.	0.28

### ***Data and Psychometric Diagnostics***

Of the 1,720 people who provided data for this scale, 52 had maximum extreme scores (i.e., selected the most positive response option for all items), and 13 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### **Reliability Coefficients:**

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.89

### **Data Fit to Model and Dimensionality:**

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 54.5% of the variance in the data (of the 54.4% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### **Item Fit:**

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?		
[AA.17] Asked questions in class.	1.11	1.06
[AA.18] Participated in class discussions.	0.87	0.86
[AA.19] Worked on reading, writing, and/or problem-solving assignments during class.	0.90	0.89
[AA.20] Discussed complex topics with other students during class.	0.76	0.76
[AA.21] Provided feedback on other students' work during class.	0.90	0.86
[AA.22] Worked with other students on group projects during class.	0.89	0.89
[AA.23] Presented my work, or work done as part of a group, to the class.	0.92	0.91
[AA.24] Worked with classmates outside of class on class assignments, homework, or projects.	1.34	1.34
[AA.25] Worked to meet the high expectations of my instructors.	1.08	1.17
[AA.26] Developed friendships with classmates.	1.07	1.06
[AA.27] Took advantage of opportunities to interact with faculty outside of the classroom.	1.22	1.18

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?					
[AA.17] Asked questions in class.	-2.09	-0.43	0.30	1.19	2.76
[AA.18] Participated in class discussions.	-2.20	-0.80	0.22	1.11	2.54
[AA.19] Worked on reading, writing, and/or problem-solving assignments during class.	-2.29	-0.82	0.20	1.06	2.54
[AA.20] Discussed complex topics with other students during class.	2.41	-0.79	0.22	1.15	2.69
[AA.21] Provided feedback on other students' work during class.	-2.31	-0.60	0.41	1.23	3.21
[AA.22] Worked with other students on group projects during class.	-2.08	-0.71	0.30	1.20	2.86
[AA.23] Presented my work, or work done as part of a group, to the class.	-2.01	-0.75	0.26	1.21	2.85
[AA.24] Worked with classmates outside of class on class assignments, homework, or projects.	-1.56	-0.26	0.35	1.10	2.81
[AA.25] Worked to meet the high expectations of my instructors.	-2.14	-0.82	0.10	0.90	2.07
[AA.26] Developed friendships with classmates.	-1.92	-0.77	0.28	1.08	2.39
[AA.27] Took advantage of opportunities to interact with faculty outside of the classroom.	-2.01	-0.36	0.48	1.11	2.98

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.75
2	-1.07
3	0.87
4	1.95

Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

***Summary***

This scale functions well as a measure of Engagement—Shared Academic Courses. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options.

## 30. Faculty Course-Related Interaction, Non Learning Community [T#FCRNLC]

### *Items*

This scale consists of five items that ask students about their interactions with faculty at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a faculty member during your {year} year at {INSTITUTION}?					
[E.25] Visited informally before or after class?	✓	✓	✓	✓	✓
[E.26] Met in his or her office about a course?	✓	✓	✓	✓	✓
[E.27] Asked (or e-mailed) him or her for information about a course?	✓	✓	✓	✓	✓
[E.47] Discussed course topics, ideas, or concepts outside of class?	X	✓	✓	✓	✓
[E.48] Discussed your academic performance?	X	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

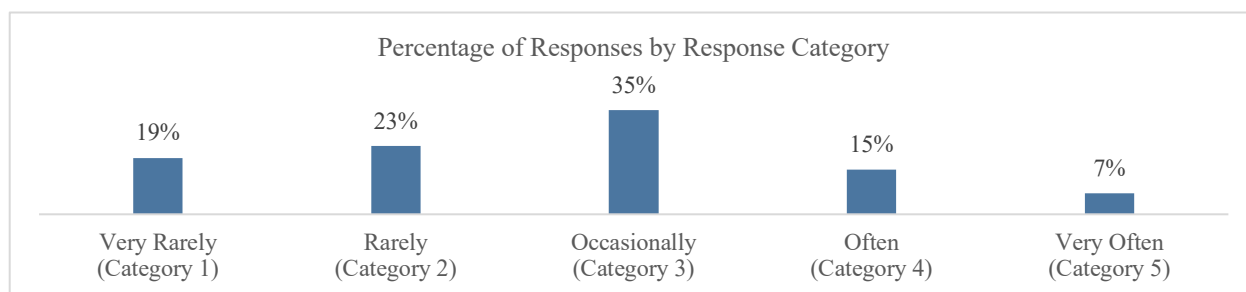
Item difficulties for this scale range from -1.20 to 0.73.

Item	Item Difficulty (logits)
How often have you done the following with a faculty member during your {year} year at {INSTITUTION}?	
[E.27] Asked (or e-mailed) him or her for information about a course?	-1.20
[E.48] Discussed your academic performance?	0.02
[E.25] Visited informally before or after class?	0.18
[E.26] Met in his or her office about a course?	0.27
[E.47] Discussed course topics, ideas, or concepts outside of class?	0.73

### *Data and Psychometric Diagnostics*

Of the 2,627 people who provided data for this scale, 43 had maximum extreme scores (i.e., selected the most positive response option for all items), and 59 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.85
- Rasch reliability: 0.78

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 58.5% of the variance in the data (of the 57.7% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a faculty member during your {year} year at {INSTITUTION}?		
[E.25] Visited informally before or after class?	0.96	0.96
[E.26] Met in his or her office about a course?	0.80	0.82
[E.27] Asked (or e-mailed) him or her for information about a course?	1.14	1.11
[E.47] Discussed course topics, ideas, or concepts outside of class?	0.97	1.01
[E.48] Discussed your academic performance?	0.97	0.97

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following with a faculty member during your {year} year at {INSTITUTION}?					
[E.25] Visited informally before or after class?	-2.61	-1.22	-0.27	0.77	2.78
[E.26] Met in his or her office about a course?	-2.51	-1.25	-0.21	1.06	3.18
[E.27] Asked (or e-mailed) him or her for information about a course?	-3.59	-1.94	-0.95	-0.01	1.51
[E.47] Discussed course topics, ideas, or concepts outside of class?	-2.18	-0.93	0.04	1.08	3.08
[E.48] Discussed your academic performance?	-2.62	-1.30	-0.39	0.76	2.64



Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart, or nearly one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.17
2	-1.27
3	1.02
4	2.41

Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

***Summary***

This scale functions well as a measure of Faculty Course-Related Interaction, Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options.

## 31. Faculty Course-Related Interaction, Learning Community About Non Learning Community [T#FCRLCNLC]

### *Items*

This scale consists of five items that ask students about their interactions with faculty at their institution who are not part of their learning community. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
[E.1] Visited informally before or after class.	✓	✓	✓	✓
[E.2] Met in his or her office about a course.	✓	✓	✓	✓
[E.3] Asked (or e-mailed) him or her for information about a course.	✓	✓	✓	✓
[E.39] Discussed course topics, ideas, or concepts outside of class.	X	✓	✓	✓
[E.40] Discussed your academic performance.	X	✓	✓	✓

### Item Difficulties and Hierarchy:

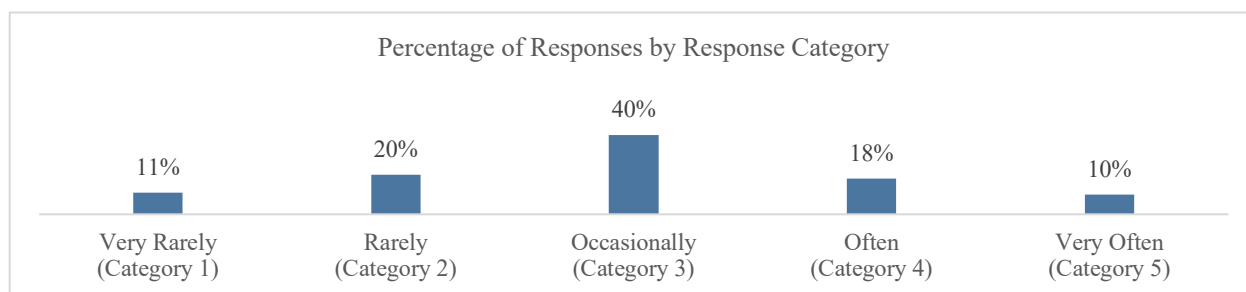
Item difficulties for this scale range from -0.80 to 0.66.

Item Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	Item Difficulty (logits)
[E.3] Asked (or e-mailed) him or her for information about a course.	-0.80
[E.40] Discussed your academic performance.	-0.06
[E.1] Visited informally before or after class.	0.10
[E.2] Met in his or her office about a course.	0.10
[E.39] Discussed course topics, ideas, or concepts outside of class.	0.66

### *Data and Psychometric Diagnostics*

Of the 1,719 people who provided data for this scale, 59 had maximum extreme scores (i.e., selected the most positive response option for all items), and 27 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.80
- Rasch reliability: 0.77

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 55.3% of the variance in the data (of the 53.8% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.		
[E.1] Visited informally before or after class.	0.87	0.87
[E.2] Met in his or her office about a course.	0.87	0.87
[E.3] Asked (or e-mailed) him or her for information about a course.	1.02	1.02
[E.39] Discussed course topics, ideas, or concepts outside of class.	1.05	1.06
[E.40] Discussed your academic performance.	0.99	0.99

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.					
[E.1] Visited informally before or after class.	-2.41	-1.15	-0.08	1.10	3.26
[E.2] Met in his or her office about a course.	-2.37	-1.25	-0.07	1.15	3.17
[E.3] Asked (or e-mailed) him or her for information about a course.	-3.14	-1.54	-0.60	0.44	2.33
[E.39] Discussed course topics, ideas, or concepts outside of class.	-1.86	-0.88	0.14	1.23	3.37
[E.40] Discussed your academic performance.	-2.47	-1.16	-0.25	0.90	2.88

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart, or nearly one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.28
2	-1.35
3	1.17
4	2.46

Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item	Evidence of DIF (and item difficulty) Between...	Interpretation
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.		
[E.40] Discussed your academic performance.	White (0.07) versus Black (-0.49)	Black students found this item easier to endorse compared with White students with comparable ability on other items.

**Summary**

This scale functions well as a measure of Faculty Course-Related Interaction, Learning Community About Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. Rasch reliability of 0.77 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 32. Faculty Course-Related Interaction, Learning Community [T#FCRLC]

### *Items*

This scale consists of five items that ask students about their interactions with faculty in their learning community. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?				
[E.13] Visited informally before or after class?	✓	✓	✓	✓
[E.14] Met in his or her office about a course?	✓	✓	✓	✓
[E.15] Asked (or e-mailed) him or her for information about a course?	✓	✓	✓	✓
[E.43] Discussed course topics, ideas, or concepts outside of class?	X	✓	✓	✓
[E.44] Discussed your academic performance?	X	✓	✓	✓

### Item Difficulties and Hierarchy:

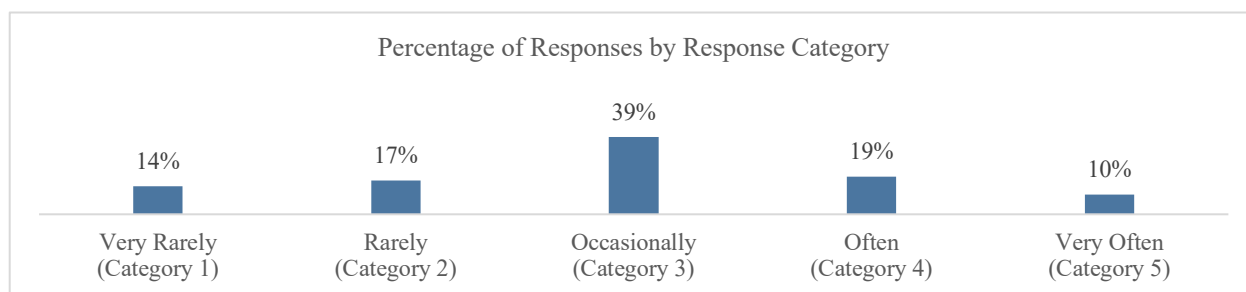
Item difficulties for this scale range from -0.53 to 0.39.

Item	Item Difficulty (logits)
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?	
[E.44] Discussed your academic performance?	-0.53
[E.15] Asked (or e-mailed) him or her for information about a course?	-0.29
[E.14] Met in his or her office about a course?	0.21
[E.13] Visited informally before or after class?	0.23
[E.43] Discussed course topics, ideas, or concepts outside of class?	0.39

### *Data and Psychometric Diagnostics*

Of the 1,719 people who provided data for this scale, 57 had maximum extreme scores (i.e., selected the most positive response option for all items), and 66 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.88
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 56.5% of the variance in the data (of the 55.9% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?		
[E.13] Visited informally before or after class?	0.97	0.97
[E.14] Met in his or her office about a course?	0.82	0.82
[E.15] Asked (or e-mailed) him or her for information about a course?	0.99	1.00
[E.43] Discussed course topics, ideas, or concepts outside of class?	1.00	1.02
[E.44] Discussed your academic performance?	1.14	1.15

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?					
[E.13] Visited informally before or after class?	-2.71	-1.18	-0.03	1.25	3.28
[E.14] Met in his or her office about a course?	-2.86	-1.13	-0.01	1.44	3.40
[E.15] Asked (or e-mailed) him or her for information about a course?	-3.18	-1.44	-0.29	0.88	2.77
[E.43] Discussed course topics, ideas, or concepts outside of class?	-2.65	-1.04	0.05	1.23	3.13
[E.44] Discussed your academic performance?	-3.31	-1.70	-0.43	0.77	2.23

**Rating Scale Functioning:**

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.21
2	-1.47
3	1.11
4	2.57

**Differential Item Functioning:**

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

***Summary***

This scale functions well as a measure of Faculty Course-Related Interaction, Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

### 33. Faculty Noncourse-Related Interaction, Non Learning Community [T#FNCRNLC]

#### *Items*

This scale consists of six items that ask students about their interactions with faculty at their institution about nonacademic matters. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more nonacademic interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

#### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?					
[E.28] Worked on an instructor's research project with him or her?	✓	✓	✓	✓	✓
[E.29] Discussed personal problems or concerns?	✓	✓	✓	✓	✓
[E.30] Discussed career plans and ambitions?	✓	✓	✓	✓	✓
[E.31] Socialized informally outside of class time?	✓	✓	✓	✓	✓
[E.49] Gotten to know each other as people outside the classroom?	X	✓	✓	✓	✓
[E.50] Discussed issues of interest or importance to you outside the classroom?	X	✓	✓	✓	✓

#### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.07 to 0.83.

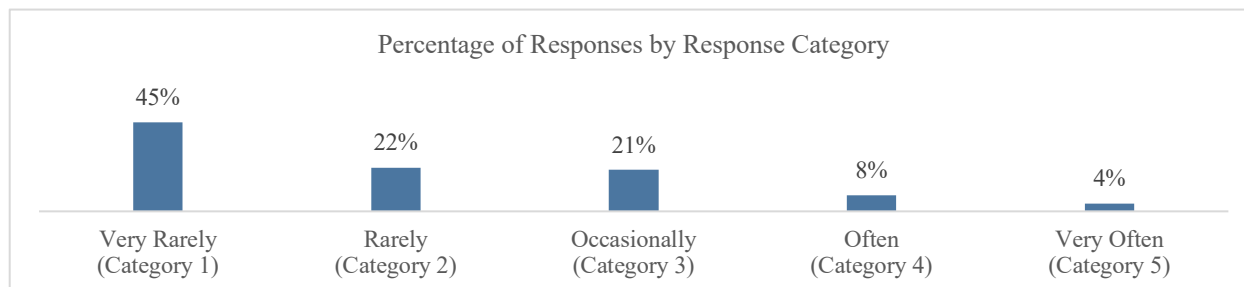
Item	Item Difficulty (logits)
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	
[E.30] Discussed career plans and ambitions?	-1.07
[E.50] Discussed issues of interest or importance to you outside the classroom?	-0.10
[E.31] Socialized informally outside of class time?	0.03
[E.49] Gotten to know each other as people outside the classroom?	0.05
[E.29] Discussed personal problems or concerns?	0.26
[E.28] Worked on an instructor's research project with him or her?	0.83



### ***Data and Psychometric Diagnostics***

Of the 2,627 people who provided data for this scale, 25 had maximum extreme scores (i.e., selected the most positive response option for all items), and 310 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.87
- Rasch reliability: 0.72

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 57.1% of the variance in the data (of the 57.0% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?		
[E.28] Worked on an instructor's research project with him or her?	1.43	1.56
[E.29] Discussed personal problems or concerns?	0.98	0.99
[E.30] Discussed career plans and ambitions?	0.96	0.95
[E.31] Socialized informally outside of class time?	0.87	0.90
[E.49] Gotten to know each other as people outside the classroom?	0.82	0.89
[E.50] Discussed issues of interest or importance to you outside the classroom?	0.70	0.72

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.28] Worked on an instructor's research project with him or her?	-2.31	-1.14	-0.35	0.29	1.83
[E.29] Discussed personal problems or concerns?	-2.77	-1.33	-0.43	0.45	1.96
[E.30] Discussed career plans and ambitions?	-3.65	-2.19	-1.20	-0.29	1.13
[E.31] Socialized informally outside of class time?	-2.91	-1.47	-0.57	0.30	1.98
[E.49] Gotten to know each other as people outside the classroom?	-2.91	-1.36	-0.45	0.24	1.79
[E.50] Discussed issues of interest or importance to you outside the classroom?	-3.08	-1.53	-0.53	0.35	2.25

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.49
2	-0.96
3	0.77
4	1.67

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	Evidence of DIF (and item difficulty) Between...	Interpretation
[E.28] Worked on an instructor's research project with him or her?	White (0.97) versus Asian (0.40)	Asian students found this item easier to endorse compared with White students with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Faculty Noncourse-Related Interaction, Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options. Rasch reliability of 0.72 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 34. Faculty Noncourse-Related Interaction, Learning Community About Non Learning Community [T#FNCRLC]

### Items

This scale consists of six items that ask students about their interactions with faculty at their institution outside of their learning community about nonacademic matters. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more nonacademic interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?				
[E.4] Worked on an instructor's research project with him or her?	✓	✓	✓	✓
[E.5] Discussed personal problems or concerns?	✓	✓	✓	✓
[E.6] Discussed career plans and ambitions?	✓	✓	✓	✓
[E.7] Socialized informally outside of class time?	✓	✓	✓	✓
[E.41] Gotten to know each other as people outside the classroom?	X	✓	✓	✓
[E.42] Discussed issues of interest or importance to you outside the classroom?	X	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.82 to 0.86.

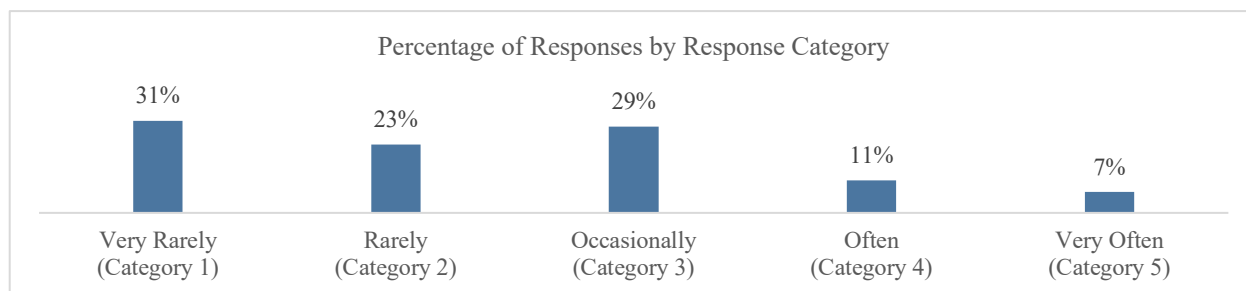
Item	Item Difficulty (logits)
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	
[E.6] Discussed career plans and ambitions?	-0.82
[E.42] Discussed issues of interest or importance to you outside the classroom?	-0.15
[E.41] Gotten to know each other as people outside the classroom?	-0.14
[E.7] Socialized informally outside of class time?	0.01
[E.5] Discussed personal problems or concerns?	0.22
[E.4] Worked on an instructor's research project with him or her?	0.86

### Data and Psychometric Diagnostics

Of the 1,719 people who provided data for this scale, 33 had maximum extreme scores (i.e., selected the most positive response option for all items), and 126 had minimum extreme scores

(i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.88
- Rasch reliability: 0.79

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 57.5% of the variance in the data (of the 56.5% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?		
[E.4] Worked on an instructor's research project with him or her?	1.47	1.51
[E.5] Discussed personal problems or concerns?	0.92	0.93
[E.6] Discussed career plans and ambitions?	0.98	0.95
[E.7] Socialized informally outside of class time?	0.82	0.84
[E.41] Gotten to know each other as people outside the classroom?	0.83	0.84
[E.42] Discussed issues of interest or importance to you outside the classroom?	0.78	0.79

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.4] Worked on an instructor's research project with him or her?	-1.99	-1.04	-0.15	0.79	2.81
[E.5] Discussed personal problems or concerns?	-2.65	-1.31	-0.27	0.58	2.71
[E.6] Discussed career plans and ambitions?	-3.61	-1.85	-0.92	0.17	1.63
[E.7] Socialized informally outside of class time?	-2.86	-1.35	-0.43	0.60	2.63
[E.41] Gotten to know each other as people outside the classroom?	-2.98	-1.45	-0.50	0.53	2.37
[E.42] Discussed issues of interest or importance to you outside the classroom?	-3.03	-1.51	-0.47	0.65	2.46

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.73
2	-1.16
3	1.00
4	1.89

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item How often have you done the following with a faculty member outside of your {COMMUNITY} during your {year} year at {INSTITUTION}?	Evidence of DIF (and item difficulty) Between...	Interpretation
[E.4] Worked on an instructor's research project with him or her?	White (1.08) versus Asian (0.43)	Asian students found this item easier to endorse compared with White students with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Faculty Noncourse-Related Interaction, Learning Community About Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with appropriate response options.

## 35. Faculty Noncourse-Related Interaction, Learning Community [T#FNCRLC]

### *Items*

This scale consists of six items that ask students about their interactions with faculty associated with their learning community about nonacademic matters. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more nonacademic interaction.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?				
[E.16] Worked on an instructor's research project with him or her?	✓	✓	✓	✓
[E.17] Discussed personal problems or concerns?	✓	✓	✓	✓
[E.18] Discussed career plans and ambitions?	✓	✓	✓	✓
[E.19] Socialized informally outside of class time?	✓	✓	✓	✓
[E.45] Gotten to know each other as people outside the classroom?	X	✓	✓	✓
[E.46] Discussed issues of interest or importance to you outside the classroom?	X	✓	✓	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.74 to 1.15.

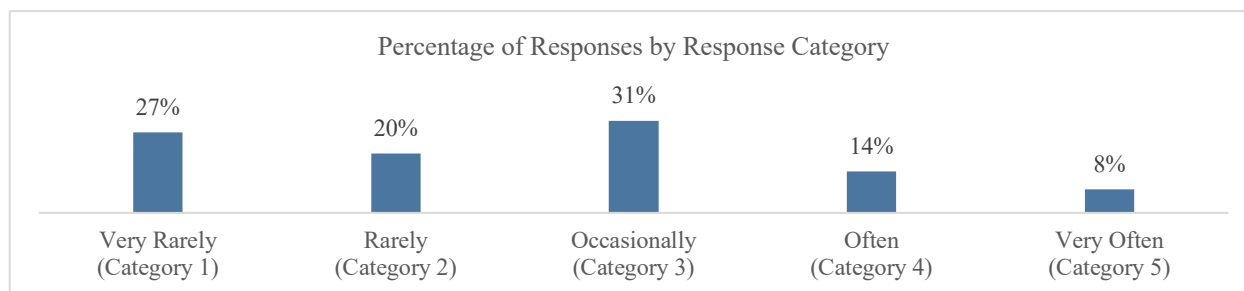
Item	Item Difficulty (logits)
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?	
[E.18] Discussed career plans and ambitions?	-0.74
[E.46] Discussed issues of interest or importance to you outside the classroom?	-0.24
[E.45] Gotten to know each other as people outside the classroom?	-0.13
[E.19] Socialized informally outside of class time?	-0.02
[E.17] Discussed personal problems or concerns?	-0.01
[E.16] Worked on an instructor's research project with him or her?	1.15

### *Data and Psychometric Diagnostics*

Of the 1,719 people who provided data for this scale, 39 had maximum extreme scores (i.e., selected the most positive response option for all items), and 124 had minimum extreme scores

(i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.89
- Rasch reliability: 0.81

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 58.2% of the variance in the data (of the 57.6% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?		
[E.16] Worked on an instructor's research project with him or her?	1.56	1.68
[E.17] Discussed personal problems or concerns?	0.88	0.89
[E.18] Discussed career plans and ambitions?	0.97	0.96
[E.19] Socialized informally outside of class time?	0.85	0.84
[E.45] Gotten to know each other as people outside the classroom?	0.82	0.83
[E.46] Discussed issues of interest or importance to you outside the classroom?	0.77	0.78

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.16] Worked on an instructor's research project with him or her?	-1.80	-0.70	0.03	1.02	3.17
[E.17] Discussed personal problems or concerns?	-2.93	-1.39	-0.30	0.69	2.74
[E.18] Discussed career plans and ambitions?	-3.73	-1.86	-0.70	0.24	1.94
[E.19] Socialized informally outside of class time?	-3.08	-1.36	-0.30	0.79	2.57
[E.45] Gotten to know each other as people outside the classroom?	-3.19	-1.41	-0.35	0.64	2.55
[E.46] Discussed issues of interest or importance to you outside the classroom?	-3.42	-1.48	-0.38	0.59	2.62

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-1.85
2	-1.21
3	0.96
4	2.10

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item How often have you done the following with a faculty member connected to your {COMMUNITY} during your {year} year at {INSTITUTION}?	Evidence of DIF (and item difficulty) Between...	Interpretation
[E.16] Worked on an instructor's research project with him or her?	White (1.52) versus Black (1.00), Hispanic (0.91), and Asian (0.74)	White students found this item more difficult to endorse compared with their peers of other races/ethnicities with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Faculty Noncourse-Related Interaction, Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.



## 36. Faculty Encouragement, Non Learning Community [T#FENLC]

### *Items*

This scale consists of five items that ask students about their perceptions of faculty encouragement with faculty at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more encouragement perceived.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

This scale was only offered one time; therefore, no changes were made over time.

Item	Cohort 2015, T1
Rate how strongly you agree with each of the following statements about a faculty member during your first year at {INSTITUTION}.	
[E.34] Faculty provided me with feedback that helped me assess my progress in class.	✓
[E.35] Faculty encouraged me to ask questions.	✓
[E.36] Faculty encouraged me to participate in discussions.	✓
[E.37] Faculty believed in my potential to succeed academically.	✓
[E.38] At least one faculty member has taken an interest in my development.	✓

### Item Difficulties and Hierarchy:

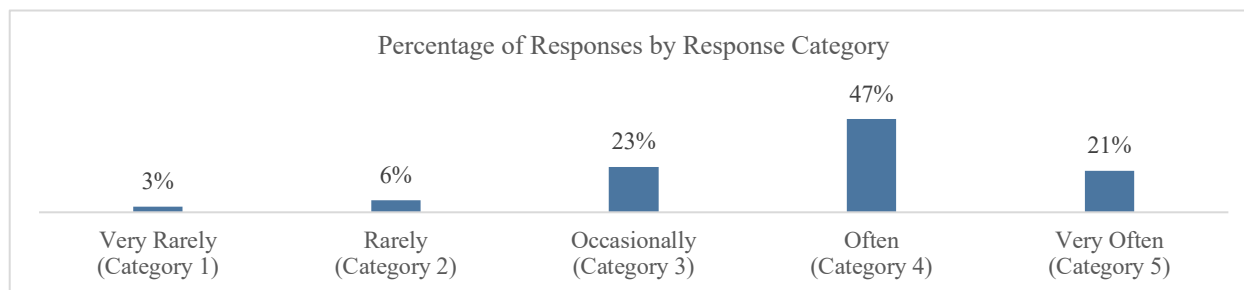
Item difficulties for this scale range from -0.66 to 0.68.

Item	Item Difficulty (logits)
Rate how strongly you agree with each of the following statements about a faculty member during your first year at {INSTITUTION}.	
[E.36] Faculty encouraged me to participate in discussions.	-0.66
[E.35] Faculty encouraged me to ask questions.	-0.45
[E.37] Faculty believed in my potential to succeed academically.	-0.02
[E.38] At least one faculty member has taken an interest in my development.	0.45
[E.34] Faculty provided me with feedback that helped me assess my progress in class.	0.68

### *Data and Psychometric Diagnostics*

Of the 459 people who provided data for this scale, 20 had maximum extreme scores (i.e., selected the most positive response option for all items), and 4 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.87
- Rasch reliability: 0.81

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 60.2% of the variance in the data (of the 60.2% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Rate how strongly you agree with each of the following statements about a faculty member during your first year at {INSTITUTION}.		
[E.34] Faculty provided me with feedback that helped me assess my progress in class.	1.02	0.99
[E.35] Faculty encouraged me to ask questions.	0.81	0.83
[E.36] Faculty encouraged me to participate in discussions.	0.92	0.91
[E.37] Faculty believed in my potential to succeed academically.	0.79	0.79
[E.38] At least one faculty member has taken an interest in my development.	1.43	1.41

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Rate how strongly you agree with each of the following statements about a faculty member during your first year at {INSTITUTION}.					
[E.34] Faculty provided me with feedback that helped me assess my progress in class.	-3.19	-0.72	0.92	2.53	5.04
[E.35] Faculty encouraged me to ask questions.	-4.10	-1.03	0.10	1.75	4.50
[E.36] Faculty encouraged me to participate in discussions.	-4.31	-1.43	-0.13	1.82	4.11

Item Rate how strongly you agree with each of the following statements about a faculty member during your first year at {INSTITUTION}.	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.37] Faculty believed in my potential to succeed academically.	-4.10	-1.15	0.35	2.09	4.69
[E.38] At least one faculty member has taken an interest in my development.	-2.86	-0.54	0.95	2.19	4.51

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.92
2	-1.72
3	0.57
4	4.07

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender or race/ethnicity.

### ***Summary***

This scale functions well as a measure of Faculty Encouragement, Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options.

## 37. Faculty Encouragement, Learning Community About Non Learning Community [T#FELCNLC]

### *Items*

This scale consists of five items that ask students about their perceptions of faculty encouragement with faculty at their institution, outside of their learning community. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more encouragement perceived.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

This scale was only offered one time; therefore, no changes were made over time.

Item	Cohort 2015, T1
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	
[E.8] Faculty provided me with feedback that helped me assess my progress in class.	✓
[E.9] Faculty encouraged me to ask questions.	✓
[E.10] Faculty encouraged me to participate in discussions.	✓
[E.11] Faculty believed in my potential to succeed academically.	✓
[E.12] At least one faculty member has taken an interest in my development.	✓

### Item Difficulties and Hierarchy:

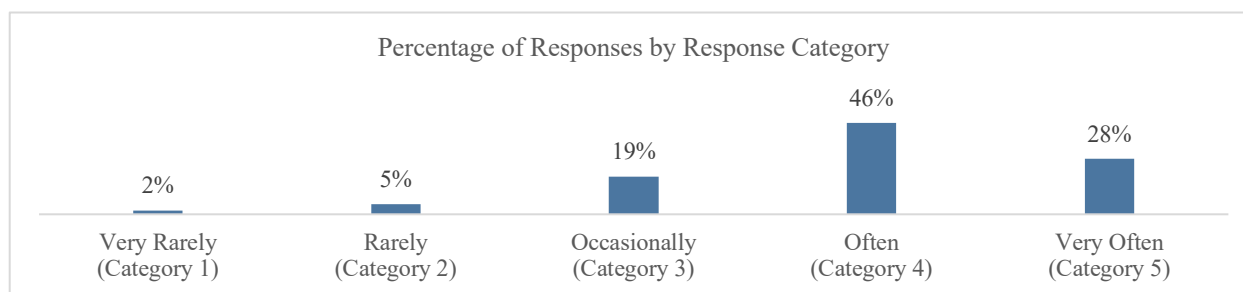
Item difficulties for this scale range from -0.42 to 0.59.

Item	Item Difficulty (logits)
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	
[E.10] Faculty encouraged me to participate in discussions.	-0.42
[E.11] Faculty believed in my potential to succeed academically.	-0.33
[E.9] Faculty encouraged me to ask questions.	-0.25
[E.12] At least one faculty member has taken an interest in my development.	0.41
[E.8] Faculty provided me with feedback that helped me assess my progress in class.	0.59

### *Data and Psychometric Diagnostics*

Of the 463 people who provided data for this scale, 44 had maximum extreme scores (i.e., selected the most positive response option for all items), and 4 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.85
- Rasch reliability: 0.78

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 52.1% of the variance in the data (of the 52.0% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.		
[E.8] Faculty provided me with feedback that helped me assess my progress in class.	0.99	1.01
[E.9] Faculty encouraged me to ask questions.	0.83	0.81
[E.10] Faculty encouraged me to participate in discussions.	0.89	0.88
[E.11] Faculty believed in my potential to succeed academically.	0.62	0.62
[E.12] At least one faculty member has taken an interest in my development.	1.58	1.58

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.					
[E.8] Faculty provided me with feedback that helped me assess my progress in class.	-2.41	-0.31	0.79	2.43	5.15
[E.9] Faculty encouraged me to ask questions.	-4.09	-0.50	0.03	1.97	4.74
[E.10] Faculty encouraged me to participate in discussions.	-4.54	-1.00	0.22	1.84	4.50

Item Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.11] Faculty believed in my potential to succeed academically.	-5.57	-1.55	0.14	1.89	4.73
[E.12] At least one faculty member has taken an interest in my development.	-1.78	-0.27	1.17	2.00	4.41

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.71
2	-1.47
3	0.40
4	3.77

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item Rate how strongly you agree with each of the following statements about a faculty member outside of your {COMMUNITY} during your first year at {INSTITUTION}.	Evidence of DIF (and item difficulty) Between...	Interpretation
[E.10] Faculty encouraged me to participate in discussions.	Black (0.26) versus Asian (-1.00) and White (-0.48)	Black students found this item more difficult to endorse compared with Asian and White students with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Faculty Encouragement, Learning Community About Non Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. Rasch reliability of 0.78 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 38. Faculty Encouragement, Learning Community [T#FELC]

### *Items*

This scale consists of five items that ask students about their perceptions of faculty encouragement with faculty at their institution, within their learning community. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more encouragement perceived.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

This scale was only offered one time; therefore, no changes were made over time.

Item	Cohort 2015, T1
Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.	
[E.20] Faculty provided me with feedback that helped me assess my progress in class.	✓
[E.21] Faculty encouraged me to ask questions.	✓
[E.22] Faculty encouraged me to participate in discussions.	✓
[E.23] Faculty believed in my potential to succeed academically.	✓
[E.24] At least one faculty member has taken an interest in my development.	✓

### Item Difficulties and Hierarchy:

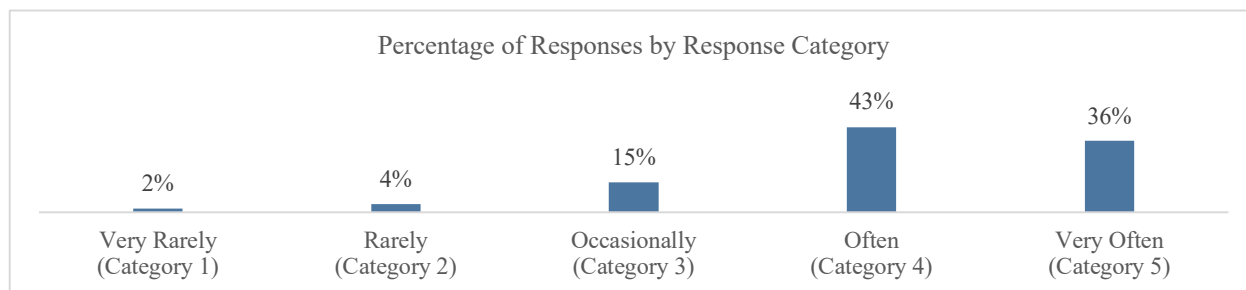
Item difficulties for this scale range from -0.60 to 0.61.

Item	Item Difficulty (logits)
Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.	
[E.20] Faculty provided me with feedback that helped me assess my progress in class.	0.61
[E.21] Faculty encouraged me to ask questions.	0.12
[E.22] Faculty encouraged me to participate in discussions.	-0.31
[E.23] Faculty believed in my potential to succeed academically.	-0.60
[E.24] At least one faculty member has taken an interest in my development.	0.18

### *Data and Psychometric Diagnostics*

Of the 463 people who provided data for this scale, 64 had maximum extreme scores (i.e., selected the most positive response option for all items), and 3 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents. Rasch reliability of 0.77 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

Across all items in this scale, the percentage of responses in each category is as follows:



### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.88
- Rasch reliability: 0.77

### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 57.8% of the variance in the data (of the 58.3% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.		
[E.20] Faculty provided me with feedback that helped me assess my progress in class.	1.01	0.99
[E.21] Faculty encouraged me to ask questions.	0.99	0.93
[E.22] Faculty encouraged me to participate in discussions.	0.87	0.89
[E.23] Faculty believed in my potential to succeed academically.	0.74	0.80
[E.24] At least one faculty member has taken an interest in my development.	1.33	1.41

### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.					
[E.20] Faculty provided me with feedback that helped me assess my progress in class.	-2.48	-0.66	0.84	2.65	5.23
[E.21] Faculty encouraged me to ask questions.	-2.37	-1.00	0.61	2.16	5.07
[E.22] Faculty encouraged me to participate in discussions.	-2.52	-1.49	0.16	1.96	4.74



Item Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
[E.23] Faculty believed in my potential to succeed academically.	-3.89	-1.44	0.01	1.74	4.68
[E.24] At least one faculty member has taken an interest in my development.	-3.03	-0.14	.055	2.20	4.94

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart or nearly one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.32
2	-1.71
3	0.30
4	3.72

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item Rate how strongly you agree with each of the following statements about a faculty member connected to {COMMUNITY} during your first year at {INSTITUTION}.	Evidence of DIF (and item difficulty) Between...	Interpretation
[E.24] At least one faculty member has taken an interest in my development.	Black (-0.42) versus White (0.30)	Black students found this item easier to endorse compared with White students with comparable ability on other items.
[E.23] Faculty believed in my potential to succeed academically.	Hispanic (-0.31) versus White (-0.81)	White students found this item easier to endorse compared with Hispanic students with comparable ability on other items.

### **Summary**

This scale functions well as a measure of Faculty Encouragement, Learning Community. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. Rasch reliability of 0.77 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 39. Educational Resiliency [T#RESILIENCY]

### *Items*

This scale consists of seven items that ask students about their perceptions of their own resiliency. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more nonacademic interaction.

The response options for each item in this scale are:

Strongly Disagree, Disagree, Agree, Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

Item	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements.				
[V.1] I am able to adapt to change in educational situations or settings.	✓	✓	✓	✓
[V.2] I can deal with whatever comes in educational situations or settings.	✓	✓	✓	✓
[V.4] Coping with stress can strengthen me in educational situations or settings.	✓	✓	✓	✓
[V.6] I can achieve my educational goals despite obstacles.	✓	✓	✓	✓
[V.7] I can stay focused on my assignments and coursework under pressure.	✓	✓	✓	✓
[V.8] I am not easily discouraged by failure in my courses.	✓	✓	✓	✓
[V.10] I can handle unpleasant feelings related to my educational experiences.	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

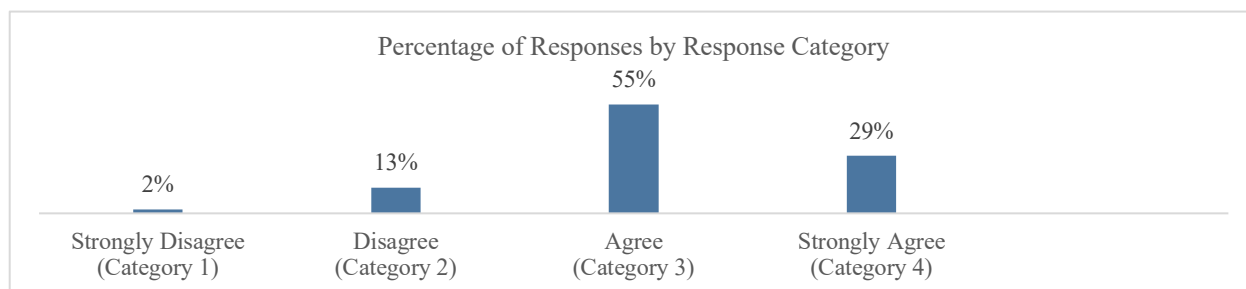
Item difficulties for this scale range from -0.91 to 1.60.

Item	Item Difficulty (logits)
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements.	
[V.6] I can achieve my educational goals despite obstacles.	-0.91
[V.1] I am able to adapt to change in educational situations or settings.	-0.71
[V.2] I can deal with whatever comes in educational situations or settings.	-0.31
[V.4] Coping with stress can strengthen me in educational situations or settings.	-0.08
[V.7] I can stay focused on my assignments and coursework under pressure.	-0.01
[V.10] I can handle unpleasant feelings related to my educational experiences.	0.43
[V.8] I am not easily discouraged by failure in my courses.	1.60

### ***Data and Psychometric Diagnostics***

Of the 3,423 people who provided data for this scale, 255 had maximum extreme scores (i.e., selected the most positive response option for all items), and 4 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.83
- Rasch reliability: 0.78

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 47.7% of the variance in the data (of the 47.5% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements.		
[V.1] I am able to adapt to change in educational situations or settings.	0.85	0.85
[V.2] I can deal with whatever comes in educational situations or settings.	0.77	0.77
[V.4] Coping with stress can strengthen me in educational situations or settings.	1.12	1.16
[V.6] I can achieve my educational goals despite obstacles.	0.83	0.83
[V.7] I can stay focused on my assignments and coursework under pressure.	0.99	1.00
[V.8] I am not easily discouraged by failure in my courses.	1.49	1.34
[V.10] I can handle unpleasant feelings related to my educational experiences.	0.88	0.89

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements.	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
[V.1] I am able to adapt to change in educational situations or settings.	-1.62	-0.52	1.17	4.00
[V.2] I can deal with whatever comes in educational situations or settings.	-1.83	-0.46	1.35	4.31
[V.4] Coping with stress can strengthen me in educational situations or settings.	-1.12	0.11	1.43	4.12
[V.6] I can achieve my educational goals despite obstacles.	-1.98	-0.62	1.07	3.86
[V.7] I can stay focused on my assignments and coursework under pressure.	-1.47	-0.01	1.46	4.25
[V.8] I am not easily discouraged by failure in my courses.	-0.15	0.90	2.06	4.88
[V.10] I can handle unpleasant feelings related to my educational experiences.	-1.32	0.14	1.66	4.74

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.64
2	-0.76
3	3.40

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender, but evidence of DIF was found by race/ethnicity as follows:

Item Thinking about yourself as a student, please indicate the extent to which you agree or disagree with the following statements.	Evidence of DIF (and item difficulty) Between...	Interpretation
[V.8] I am not easily discouraged by failure in my courses.	Asian (0.80) versus Black (1.31), Hispanic (1.34), and White (1.81)	Asian students found this item easier to endorse compared with their peers of other races/ethnicities with comparable ability on other items.
[V.6] I can achieve my educational goals despite obstacles.	Asian (-0.48) versus White (0.02)	Asian students found this item easier to endorse compared with their White peers with comparable ability on other items.

### Summary

This scale functions well as a measure of Educational Resiliency. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. All items have acceptable fit and diagnostics; however, item V.8 stands out as having the largest fit statistics and evidence of DIF (and is the only item with “not”). For future iterations of this measure, consider modifying this item with positive wording.

Additionally, Rasch reliability of 0.78 indicates that the scale reliably differentiates among people of higher and lower ability levels, but could be improved with additional items, or items of more varied difficulty.

## 40. College Knowledge [T#CK]

### Items

This scale consists of two items that ask students about their confidence in their ability to navigate specific college tasks. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of social peer interactions.

The response options for each item in this scale are:

Cannot Do This At All        Absolutely Can Do This

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:					
[M.13] Write a college-level paper.	✓	✓	✓	✓	✓
[M.15] Make the best use of faculty office hours.	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

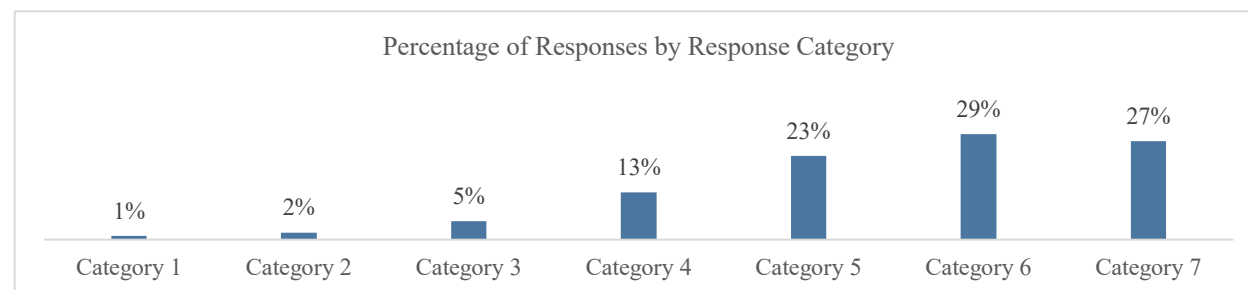
Item difficulties for this scale range from -0.55 to 0.55.

Item	Item Difficulty (logits)
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:	
[M.13] Write a college-level paper.	-0.55
[M.15] Make the best use of faculty office hours.	0.55

### Data and Psychometric Diagnostics

Of the 4,338 people who provided data for this scale, 634 had maximum extreme scores (i.e., selected the most positive response option for all items), and 6 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:



Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.58
- Rasch reliability: 0.61

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 64.5% of the variance in the data (of the 64.2% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:		
[M.13] Write a college-level paper.	1.06	1.10
[M.15] Make the best use of faculty office hours.	0.85	0.90

Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:							
[M.13] Write a college-level paper.	-3.25	-1.49	-0.84	0.06	1.04	2.33	4.23
[M.15] Make the best use of faculty office hours.	-1.98	-1.00	-0.07	0.62	1.68	3.01	5.12

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart, or nearly one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-3.18
2	-1.80
3	-0.91
4	0.26
5	1.71
6	3.92

Differential Item Functioning:

There is no evidence of DIF by gender or race/ethnicity for any item in this scale.

***Summary***

These two items function relatively well as a measure of College Knowledge, considering the number of items. As would be expected with a two-item scale, reliability coefficients are not high; however, there is evidence that the scale measures one underlying latent trait, reliably differentiates among people of high and low ability levels, and is comprised of items that each contribute meaningfully to the scale, with well-functioning response options. Adding items and/or further operationalizing or defining response categories may help to improve the scale in the future.



## 41. Staff Care and Support [T#SCS]

### *Items*

This scale consists of seven items that ask students about their experiences with the advising staff of the learning community program. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Staff Care and Support.

The response options for each item in this scale are:

- Strongly agree
- Agree
- Unsure
- Disagree
- Strongly disagree

### Changes Over Time:

There were no changes over time to the items or response options for this scale. However, one item was not asked of the 2016 cohort at T1 due to programming error in the survey instrument:

Item	Cohort 2015, T1	Cohort 2015, T2	Cohort 2016, T1	Cohort 2016, T2
Please rate your agreement with the following statements about your interactions in your {year} year with the advising staff who work in the {COMMUNITY} program. Please do not include peer mentors, or interactions you have with faculty during class, in your answers.				
[AA.34] I developed a close, personal relationship with at least one {COMMUNITY} staff member.	✓	✓	✓	✓
[AA.35] Opportunities to meet and interact informally with {COMMUNITY} staff members have helped me feel like I belong at {INSTITUTION}.	✓	✓	✓	✓
[AA.36] {COMMUNITY} staff members have high expectations of me.	✓	✓	✓	✓
[AA.37] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me to meet the high expectations they have of me here at {INSTITUTION}.	✓	✓	✓	✓
[AA.38] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had personally.	✓	✓	✓	✓
[AA.39] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had academically.	✓	✓	✓	✓
[AA.40] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me make better use of campus services not part of the {COMMUNITY}.	✓	✓	X	✓

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -1.19 to 0.58.

Item Please rate your agreement with the following statements about your interactions in your {year} year with the advising staff who work in the {COMMUNITY} program. Please do not include peer mentors, or interactions you have with faculty during class, in your answers.	Item Difficulty (logits)
[AA.36] {COMMUNITY} staff members have high expectations of me.	-1.19
[AA.35] Opportunities to meet and interact informally with {COMMUNITY} staff members have helped me feel like I belong at {INSTITUTION}.	-0.18
[AA.37] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me to meet the high expectations they have of me here at {INSTITUTION}.	-0.10
[AA.40] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me make better use of campus services not part of the {COMMUNITY}.	0.20
[AA.39] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had academically.	0.26
[AA.34] I developed a close, personal relationship with at least one {COMMUNITY} staff member.	0.42
[AA.38] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had personally.	0.58

### ***Data and Psychometric Diagnostics***

Of the 1,722 people who provided data for this scale, 158 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 2%
- Category 2 = 7%
- Category 3 = 16%
- Category 4 = 45%
- Category 5 = 30%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.89
- Rasch reliability: 0.80

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 54.8% of the variance in the data (of the 54.1% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Please rate your agreement with the following statements about your interactions in your {year} year with the advising staff who work in the {COMMUNITY} program. Please do not include peer mentors, or interactions you have with faculty during class, in your answers.		
[AA.34] I developed a close, personal relationship with at least one {COMMUNITY} staff member.	1.24	1.21
[AA.35] Opportunities to meet and interact informally with {COMMUNITY} staff members have helped me feel like I belong at {INSTITUTION}.	0.88	0.84
[AA.36] {COMMUNITY} staff members have high expectations of me.	1.14	1.22
[AA.37] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me to meet the high expectations they have of me here at {INSTITUTION}.	0.82	0.78
[AA.38] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had personally.	0.99	0.96
[AA.39] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had academically.	0.90	0.89
[AA.40] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me make better use of campus services not part of the {COMMUNITY}.	1.07	1.03

#### Ability Mean by Category:

The data for this scale follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
Please rate your agreement with the following statements about your interactions in your {year} year with the advising staff who work in the {COMMUNITY} program. Please do not include peer mentors, or interactions you have with faculty during class, in your answers.					
[AA.34] I developed a close, personal relationship with at least one {COMMUNITY} staff member.	-1.39	0.02	0.88	1.86	4.22
[AA.35] Opportunities to meet and interact informally with {COMMUNITY} staff members have helped me feel like I belong at {INSTITUTION}.	-1.87	-0.65	0.32	1.63	4.19
[AA.36] {COMMUNITY} staff members have high expectations of me.	-1.97	-0.79	-0.03	1.24	3.37
[AA.37] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me to meet the high expectations they have of me here at {INSTITUTION}.	-2.40	-0.69	0.39	1.70	4.24
[AA.38] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had personally.	-1.51	0.01	0.84	1.96	4.59
[AA.39] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me deal with struggles I have had academically.	-1.33	-0.37	0.58	1.84	4.56
[AA.40] My <b>nonclassroom</b> interactions with {COMMUNITY} staff have helped me make better use of campus services not part of the {COMMUNITY}.	-1.91	-0.25	0.78	1.79	4.38

#### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest

probability response). However, the thresholds' close distance ( $<1$  logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.47
2	-0.80
3	0.03
4	3.24

Differential Item Functioning:

For each of the eight items examined, there is no evidence of DIF by cohort or timepoint.

***Summary***

This scale functions well as a measure of Staff Care and Support. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale.

## 42. Sense of Belonging: Learning Communities [T#SOB]

### Items

This scale consists of eight items that ask students about their sense of belonging in their learning communities. Five of the eight items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and three of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more positive perceptions of sense of belonging.

The response options for each item in this scale are<sup>13</sup>:

Strongly Disagree                                Strongly Agree

### Item Changes Over Time:

Two items were added to this measure after the first survey administration as shown:

Item	Cohort 2015, T1 (4-point ROs)	Cohort 2015, T2 (7-point ROs)	Cohort 2015, T3 (7-point ROs)	Cohort 2016, T1 (7-point ROs)	Cohort 2016, T2 (7-point ROs)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:					
[R.1] I will feel like an outsider.*	✓	✓	✓	✓	✓
[R.2] I will make friends easily.	✓	✓	✓	✓	✓
[R.3] I will feel like I belong.	✓	✓	✓	✓	✓
[R.4] I will feel awkward and out of place.*	✓	✓	✓	✓	✓
[R.5] I will feel lonely.*	✓	✓	✓	✓	✓
[R.6] I expect other students will like me.	✓	✓	✓	✓	✓
[R.7] I will see myself as an important part of the {COMMUNITY} community.	X	✓	✓	✓	✓
[R.8] I will feel I am a member of the {COMMUNITY} community.	X	✓	✓	✓	✓

\*Negatively valenced item; reverse coded.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.85 to 0.39.

Item	Item Difficulty (logits)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[R.8] I will feel I am a member of the {COMMUNITY} community.	-0.85

<sup>13</sup> Students in Cohort 2015 were presented with four response options in timepoint 1. To be scaled together, data from subsequent administrations were collapsed and scale diagnostics were examined with 4-point response options. See the Methods to Scale Across Timepoints section for more information.

Item	Item Difficulty (logits)
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:	
[R.3] I will feel like I belong.	-0.32
[R.6] I expect other students will like me.	-0.04
[R.5] I will feel lonely.*	0.07
[R.7] I will see myself as an important part of the {COMMUNITY} community.	0.17
[R.1] I will feel like an outsider.*	0.22
[R.4] I will feel awkward and out of place.*	0.38
[R.2] I will make friends easily.	0.39

\*Negatively valenced item; reverse coded.

### ***Data and Psychometric Diagnostics***

Of the 1,717 people who provided data for this scale, 227 had extreme scores (i.e., selected the most positive response option for all items or the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

For consistency across administrations, response categories were collapsed to 4 points for administrations that used 7-point response options. Categories were collapsed as follows:

Original Category	Recoded Category
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 4%
- Category 2 = 16%
- Category 3 = 44%
- Category 4 = 35%

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.93
- Rasch reliability: 0.82

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 53.1% of the variance in the data (of the 51.7% of variance explained that would be expected if data are

a good fit to the model). The largest contrast accounts for 12.0% of unexplained variance, a slightly notable percentage but substantially lower than variance accounted for by the measure. Not uncommon in scales with positively and negatively valenced items, the largest contrast is defined by the positively valenced and negatively valenced items in this scale. There are no other large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the range of 0.5 to 2 MNSQ. Item-fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:		
[R.1] I will feel like an outsider.*	0.91	0.92
[R.2] I will make friends easily.	1.01	1.01
[R.3] I will feel like I belong.	0.78	0.82
[R.4] I will feel awkward and out of place.*	0.90	0.88
[R.5] I will feel lonely.*	1.00	0.95
[R.6] I expect other students will like me.	0.90	0.87
[R.7] I will see myself as an important part of the {COMMUNITY} community.	1.10	1.08
[R.8] I will feel I am a member of the {COMMUNITY} community.	1.00	1.02

\*Negatively valenced item; reverse coded.

#### Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category			
	1	2	3	4
Thinking about {COMMUNITY}, please rate the extent to which you agree or disagree with the following statements:				
[R.1] I will feel like an outsider.*	-1.30	-0.22	1.29	4.00
[R.2] I will make friends easily.	-1.53	-0.12	1.54	4.41
[R.3] I will feel like I belong.	-2.33	-0.63	1.21	4.03
[R.4] I will feel awkward and out of place.*	-1.19	-0.24	1.43	4.26
[R.5] I will feel lonely.*	-1.72	-0.21	1.30	3.90
[R.6] I expect other students will like me.	-2.27	-0.47	1.37	4.39
[R.7] I will see myself as an important part of the {COMMUNITY} community.	-1.83	-0.21	1.39	4.21
[R.8] I will feel I am a member of the {COMMUNITY} community.	-2.59	-0.78	0.85	3.65

\*Negatively valenced item; reverse coded.

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

	Positively Valenced Response Scale	Negatively Valenced Response Scale
Category	Threshold	Threshold
1	-2.56	-1.89
2	-0.47	-0.38
3	3.03	2.28

***Summary***

This scale functions well as a measure of Sense of Belonging: Learning Community. There is evidence that the scale measures one underlying latent trait, despite a small contrast attributed to the mix of positively and negatively valenced items. Furthermore, there is evidence that this scale reliably differentiates between people of varying ability levels and is comprised of items that each contribute meaningfully to the scale.



## 43. Academic Self-Efficacy: Study Skills (T1) [T1ASES]

### *Items*

This scale consists of seven items that ask students about their academic experiences with study skills at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of Academic Self-Efficacy: Study Skills.

The response options for each item in this scale are:

Cannot do this at all                                Absolutely can do this

### Changes Over Time:

There were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.65 to 0.80.

Item	Item Difficulty (logits)
Thinking about the kind of student you are at {INSTITUTION}, please rate how certain you are that you are able to do the following:	
[M.1] Finish my assignments on time.	-0.65
[M.4] Organize my schoolwork.	-0.31
[M.7] Arrange a place to study without distractions.	-0.22
[M.3] Take good notes during class.	-0.02
[M.5] Remember information presented in class.	0.19
[M.11] Have effective study skills.	0.22
[M.6] Remember information presented in textbooks.	0.80

### *Data and Psychometric Diagnostics*

Of the 5,607 people who provided data for this scale, 332 had maximum extreme scores (i.e., selected the most positive response option for all items), and 5 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 1%
- Category 2 = 2%
- Category 3 = 5%
- Category 4 = 12%
- Category 5 = 24%
- Category 6 = 30%
- Category 7 = 26%

The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.87
- Rasch reliability: 0.81

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 55.2% of the variance in the data (of the 56.0% of variance explained that would be expected if data are a good fit to the model). There are no large contrasts (factors) in the data that would indicate potential issues with multidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:		
[M.1] Finish my assignments on time.	1.25	1.34
[M.3] Take good notes during class.	0.97	1.00
[M.4] Organize my schoolwork.	0.88	0.92
[M.5] Remember information presented in class.	0.76	0.76
[M.6] Remember information presented in textbooks.	1.11	1.11
[M.7] Arrange a place to study without distractions.	1.21	1.26
[M.11] Have effective study skills.	0.80	0.80

#### Ability Mean by Category:

The data for this survey follow the expected pattern of increasing average person ability measures to increasingly positive response categories. Average person ability measures by response categories for each item are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about your experience as a student at {INSTITUTION}, please rate how certain you are that you can do the following:							
[M.1] Finish my assignments on time.	-1.66	-0.65	-0.05	0.28	0.92	1.58	2.96
[M.3] Take good notes during class.	-1.33	-0.36	-0.01	0.44	1.16	1.91	3.60
[M.4] Organize my schoolwork.	-2.65	-0.67	-0.11	0.30	0.94	1.79	3.39
[M.5] Remember information presented in class.	-2.28	-0.54	-0.07	0.49	1.26	2.11	4.13
[M.6] Remember information presented in textbooks.	-1.19	0.19	0.47	0.96	1.61	2.45	4.52
[M.7] Arrange a place to study without distractions.	-1.60	-0.51	0.09	0.49	1.10	1.83	3.24
[M.11] Have effective study skills.	-1.62	-0.49	0.07	0.51	1.24	2.12	4.04

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options (i.e., respondents do not necessarily see each as distinct). Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.36
2	-1.29
3	-0.62
4	0.13
5	1.33
6	2.81

***Summary***

This scale functions well as a measure of Academic Self-Efficacy: Study Skills. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale; however, rating scale diagnostics and response option use indicate that respondents may have been presented with more response options than needed.

## 44. Peer Collaboration in Shared Academic Courses (T1) [T1SACP]

### *Items*

This scale consists of three items that ask students about their peer collaboration experiences in shared academic courses. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of peer collaboration.

The response options for each item in this scale are:

Very Rarely, Rarely, Occasionally, Often, Very Often

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.11 to 0.19.

Item	Item Difficulty (logits)
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?	
[AA23] Presented my work, or work done as part of a group, to the class.	-0.11
[AA22] Worked with other students on group projects during class.	-0.09
[AA21] Provided feedback on other students' work during class.	0.19

### *Data and Psychometric Diagnostics*

Of the 1,718 people who provided data for this scale, 125 had maximum extreme scores (i.e., selected the most positive response option for all items), and 34 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 6%
- Category 2 = 14%
- Category 3 = 33%
- Category 4 = 31%
- Category 5 = 16%

### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.84
- Rasch reliability: 0.81

Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 66.7% of the variance in the data (of the 66.4% of variance explained that would be expected if data are a good fit to the model). There is one contrast accounting for 18.1% of unexplained variance, indicating potential issues of multidimensionality.

Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?		
[AA21] Provided feedback on other students' work during class.	1.14	1.14
[AA22] Worked with other students on group projects during class.	0.93	0.93
[AA23] Presented my work, or work done as part of a group, to the class.	0.89	0.89

Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category				
	1	2	3	4	5
How often have you done the following in your {COMMUNITY} required academic courses at {INSTITUTION}?					
[AA21] Provided feedback on other students' work during class.	-4.22	-1.78	0.34	2.34	5.24
[AA22] Worked with other students on group projects during class.	-4.08	-2.06	0.10	2.20	4.96
[AA23] Presented my work, or work done as part of a group, to the class.	-4.25	-2.14	0.07	2.30	4.89

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response) at least one logit apart. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-3.90
2	-1.72
3	1.30
4	4.31

**Summary**

This scale functions well as a measure of Peer Collaboration in Shared Academic Courses. There is evidence that the scale is comprised of items that each contribute meaningfully to the scale, and Rasch reliability of 0.81 indicates that the scale reliably differentiates among people of higher and lower ability levels. A notable secondary factor indicates that future exploration of multidimensionality may be beneficial.

## 45. Academic Validation [T#AVAL]

### *Items*

This scale consists of six items that ask students about academic support, recognition, and encouragement that they received and perceived at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of academic validation.

The response options for each item in this scale are:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale. However, cohort 2015 was not asked this scale at timepoint 1.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.53 to 0.48.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	-0.53
[OO.2] Instructors provided me with feedback that helped me judge my progress.	-0.16
[OO.1] I feel like my contributions were valued in class.	-0.06
[OO.3] Instructors were able to determine my level of understanding of course material.	0.05
[OO.6] Instructors encouraged me to meet with them after or outside of class.	0.22
[OO.5] Instructors showed concern about my progress.	0.48

### *Data and Psychometric Diagnostics*

Of the 4,685 people who provided data for this scale, 284 had maximum extreme scores (i.e., selected the most positive response option for all items), and 6 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all 6 items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 2%
- Category 2 = 3%
- Category 3 = 6%
- Category 4 = 15%
- Category 5 = 26%
- Category 6 = 29%
- Category 7 = 19%

The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.88
- Rasch reliability: 0.83

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 59.1% of the variance in the data (of the 59.1% of variance explained that would be expected if data are a good fit to the model). There is one small contrast accounting for 12.0% of unexplained variance, a much lower percentage of variance accounted for by the measure.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[OO.1] I feel like my contributions were valued in class.	1.03	0.99
[OO.2] Instructors provided me with feedback that helped me judge my progress.	0.8	0.81
[OO.3] Instructors were able to determine my level of understanding of course material.	0.78	0.75
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	0.93	0.98
[OO.5] Instructors showed concern about my progress.	1.23	1.14
[OO.6] Instructors encouraged me to meet with them after or outside of class.	1.32	1.27

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:							
[OO.1] I feel like my contributions were valued in class.	-1.83	-0.99	-0.43	0.09	0.83	1.84	3.91
[OO.2] Instructors provided me with feedback that helped me judge my progress.	-2.23	-0.97	-0.57	-0.04	0.72	1.76	3.98
[OO.3] Instructors were able to determine my level of understanding of course material.	-2.34	-0.97	-0.44	0.04	0.91	1.93	4.30
[OO.4] Instructors encouraged me to ask questions and participate in discussions.	-2.58	-1.20	-0.60	-0.17	0.52	1.51	3.42
[OO.5] Instructors showed concern about my progress.	-0.80	-0.31	-0.14	0.40	1.01	2.06	4.48
[OO.6] Instructors encouraged me to meet with them after or outside of class.	-1.11	-0.46	-0.11	0.32	0.89	1.86	4.00

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.02
2	-1.19
3	-1.06
4	-0.06
5	1.16
6	3.17

### ***Summary***

This scale functions reasonably well as a measure of Academic Validation. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale; however, rating scale diagnostics and response option use indicate that respondents may have been presented with more response options than needed.



## 46. Interpersonal Validation [T#IVAL]

### *Items*

This scale consists of six items that ask students about interpersonal support, recognition, and encouragement that they received and perceived at their institution. All items are positively valenced (i.e., where positive endorsement of an item has a positive meaning); as such, higher scores represent more positive perceptions of interpersonal validation.

The response options for each item in this scale are:

Strongly Disagree    ○    ○    ○    ○    ○    ○    ○    Strongly Agree

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale. However, cohort 2015 was not asked this scale at timepoint 1.

### Item Difficulties and Hierarchy:

Item difficulties for this scale range from -0.32 to 0.43.

Item	Item Difficulty (logits)
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:	
[OO.8] At least one faculty member has taken an interest in my development.	-0.32
[OO.7] Faculty believe in my potential to succeed academically.	-0.22
[OO.9] At least one staff member has taken an interest in my development.	-0.17
[OO.11] Faculty empower me to learn here.	0.02
[OO.12] Staff encourage me to get involved in campus activities.	0.27
[OO.10] Staff recognize my achievements.	0.43

### *Data and Psychometric Diagnostics*

Of the 4,685 people who provided data for this scale, 413 had maximum extreme scores (i.e., selected the most positive response option for all items), and 16 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all 6 items in this scale, the percentage of responses in each category is as follows:

- Category 1 = 3%
- Category 2 = 4%
- Category 3 = 6%
- Category 4 = 15%
- Category 5 = 23%
- Category 6 = 27%
- Category 7 = 23%

The minimal use of categories 1 and 2 may be an indication that more response categories than necessary were offered to respondents.

#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.91
- Rasch reliability: 0.84

#### Data Fit to Model and Dimensionality:

Data for this scale show good fit to the Rasch rating scale model. This scale accounts for 63.0% of the variance in the data (of the 63.2% of variance explained that would be expected if data are a good fit to the model). There is one small contrast accounting for 12.2% of unexplained variance.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the acceptable range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:		
[OO.7] Faculty believe in my potential to succeed academically.	0.97	0.94
[OO.8] At least one faculty member has taken an interest in my development.	1.11	1.22
[OO.9] At least one staff member has taken an interest in my development.	1.01	1.07
[OO.10] Staff recognize my achievements.	0.81	0.78
[OO.11] Faculty empower me to learn here.	0.74	0.76
[OO.12] Staff encourage me to get involved in campus activities.	1.33	1.28

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
Thinking about your experience as a student at {INSTITUTION}, please rate the extent to which you agree or disagree with the following statements:							
[OO.7] Faculty believe in my potential to succeed academically.	-2.61	-1.43	-0.71	-0.03	0.7	1.74	4.06
[OO.8] At least one faculty member has taken an interest in my development.	-2.11	-1.03	-0.47	0.00	0.62	1.51	3.50
[OO.9] At least one staff member has taken an interest in my development.	-2.16	-0.97	-0.44	0.06	0.69	1.63	3.66
[OO.10] Staff recognize my achievements.	-1.95	-0.81	-0.21	0.33	1.10	2.18	4.68
[OO.11] Faculty empower me to learn here.	-2.44	-1.15	-0.54	0.05	0.79	1.91	4.44
[OO.12] Staff encourage me to get involved in campus activities.	-1.71	-0.61	-0.16	0.33	0.97	1.95	4.40

Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are ordered, and each category peaks (i.e., each category is, at some point in the ability distribution, the highest probability response). However, the thresholds' close distance (<1 logit) for some categories indicates that there may be too many response options, and respondents do not necessarily see each as distinct. Andrich thresholds for this scale are as follows:

Category	Threshold
1	-2.00
2	-1.12
3	-1.06
4	0.10
5	1.12
6	2.96

***Summary***

This scale functions reasonably well as a measure of Interpersonal Validation. There is evidence that the scale measures one underlying latent trait, reliably differentiates among people of varying ability levels, and is comprised of items that each contribute meaningfully to the scale; however, rating scale diagnostics and response option use indicate that respondents may have been presented with more response options than needed..

## 47. Financial Stress [\*Not Delivered\*]

### Items

This scale consists of five items that ask students about aspects of financial stress. Two of the items are positively valenced (i.e., where positive endorsement of an item has a positive meaning), and three of the items are negatively valenced (i.e., where positive endorsement of an item has a negative meaning). For the creation of person-level scale scores, negatively valenced items are reverse coded such that higher scores represent more financial stress.

The response options for each item in this scale are:

Not At All    ○    ○    ○    ○    ○    ○    ○    Very Much

### Item Changes Over Time:

There were no changes over time to the items or response options for this scale

Item Please indicate the extent to which you agree with the following statements.	Cohort 2015, T1	Cohort 2015, T2	Cohort 2015, T3	Cohort 2016, T1	Cohort 2016, T2
[W.10] Financial concerns have interfered with my academic performance during my {year} year at {INSTITUTION}.	✓	✓	✓	✓	✓
[W.11] My job schedule has interfered with my academic performance during my first year at {INSTITUTION}.	✓	✓	✓	✓	✓
[W.12] I could pay for an unexpected expense of \$500.	✓	✓	✓	✓	✓
[W.13] I could pay for an unexpected expense of \$1,000.	✓	✓	✓	✓	✓
[W.14] Considering the cost of college, I think attending {INSTITUTION} is a good investment.	✓	✓	✓	✓	✓

### Item Difficulties and Hierarchy:

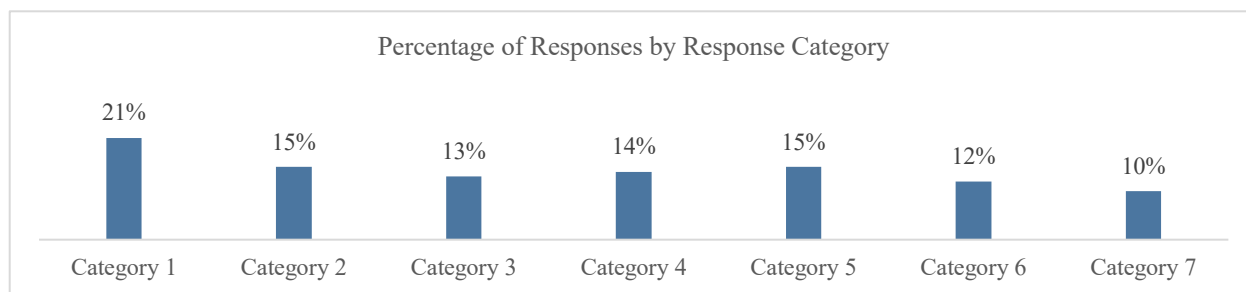
Item difficulties for this scale range from -0.56 to 0.37.

Item Please indicate the extent to which you agree with the following statements.	Item Difficulty (logits)
[W.14] Considering the cost of college, I think attending {INSTITUTION} is a good investment.	-0.56
[W.10] Financial concerns have interfered with my academic performance during my {year} year at {INSTITUTION}.	-0.02
[W.12] I could pay for an unexpected expense of \$500.	0.09
[W.11] My job schedule has interfered with my academic performance during my first year at {INSTITUTION}.	0.11
[W.13] I could pay for an unexpected expense of \$1,000.	0.37

### Data and Psychometric Diagnostics

Of the 4,346 people who provided data for this scale, 24 had maximum extreme scores (i.e., selected the most positive response option for all items), and 8 had minimum extreme scores (i.e., selected the least positive response option for all items). Our evaluation of diagnostics includes extreme and nonextreme respondents.

Across all items in this scale, the percentage of responses in each category is as follows (note the lack of a peak in the response distribution, indicating anomalies in response patterns):



#### Reliability Coefficients:

- Cronbach's  $\alpha$  (inter-item consistency): 0.38
- Rasch reliability: 0.49

#### Data Fit to Model and Dimensionality:

This scale accounts for 41.3% of the variance in the data (of the 41.1% of variance explained that would be expected if data are a good fit to the model); however, there are multiple large contrasts that suggest issues with multidimensionality in this scale, the largest of which accounts for 31% of the variance in residuals and has the power of 2.6 items. The results of this principal components analysis (PCA) provide evidence to suggest that this scale does not fit the model assumption of unidimensionality.

#### Item Fit:

All items in this scale contribute information to the construct scores and are within the ideal range of 0.5 to 2 MNSQ. Item fit statistics are as follows:

Item	Outfit MNSQ	Infit MNSQ
Please indicate the extent to which you agree with the following statements.		
[W.10] Financial concerns have interfered with my academic performance during my {year} year at {INSTITUTION}.	0.96	0.98
[W.11] My job schedule has interfered with my academic performance during my first year at {INSTITUTION}.	1.33	1.29
[W.12] I could pay for an unexpected expense of \$500.	0.75	0.82
[W.13] I could pay for an unexpected expense of \$1,000.	0.79	0.94
[W.14] Considering the cost of college, I think attending {INSTITUTION} is a good investment.	1.12	0.95

#### Ability Mean by Category:

The data for this survey follow the expected pattern of average person ability measures by response categories. Average person ability measures by response category are as follows:

Item Please indicate the extent to which you agree with the following statements.	Average Person Ability Measure (logits), by Response Category						
	1	2	3	4	5	6	7
[W.10] Financial concerns have interfered with my academic performance during my {year} year at {INSTITUTION}.	-0.51	-0.30	-0.16	-0.07	0.09	0.34	0.73
[W.11] My job schedule has interfered with my academic performance during my first year at {INSTITUTION}.	-0.45	-0.33	-0.16	-0.09	0.08	0.25	0.46
[W.12] I could pay for an unexpected expense of \$500.	-0.41	-0.23	-0.12	0.03	0.20	0.46	0.96
[W.13] I could pay for an unexpected expense of \$1,000.	-0.32	-0.06	0.14	0.27	0.43	0.67	1.10
[W.14] Considering the cost of college, I think attending {INSTITUTION} is a good investment.	-0.75	-0.45	-0.34	-0.24	-0.05	0.14	0.39

### Rating Scale Functioning:

Examination of the Andrich thresholds for this scale shows that thresholds are disordered for both rating scales. In addition, the thresholds are very close in distance, indicating that the response options are not appropriate for the items. Andrich thresholds for this scale are as follows:

	Positively Valenced Response Scale	Negatively Valenced Response Scale
Category	Threshold	Threshold
1	0.29	-0.39
2	-0.19	-0.32
3	-0.56	0.02
6	-0.15	0.29
5	0.24	0.11
6	0.38	0.28

### Differential Item Functioning:

DIF by gender and race/ethnicity was examined for each item in this scale. No evidence of DIF was found by gender or race.

### **Summary**

This scale does not function well as a measure of Financial Stress. The scale has low reliability, evidence of multidimensionality, and evidence of inappropriate response options. It is our recommendation these data should be used at the item level only, and this scale should be modified prior to any future use. Rasch scores for this measure are not included in the accompanying data file.

## Methods to Scale Across Timepoints

The TSLC follow-up survey was conducted multiple times and, in some cases, changes were made to the survey between administrations, including:

1. Change in the number of response options presented
2. Item(s) dropped from measures over time
3. Item(s) added to measures over time
4. Items reworded between collections

In this section we describe the approaches used to create scale scores that account for changes in the survey over time (and that are comparable across administrations).

### Summary of Changes:

Scale	Change Response Options	Item(s) Dropped	Item(s) Added	Items(s) Reworded
<b>Initial “Baseline” Survey Measures</b>				
1. Self-Efficacy: Expected Social Adjustment			X	X
2. Self-Efficacy: Expected Academic Adjustment			X	
3. Expected Mattering: Campus	X	X	X	X
4. Expected Sense of Belonging: Campus	X		X	
5. Expected Mattering: Learning Community				
6./7. Resiliency				X
8. Expected Sense of Belonging: Learning Communities	X		X	
9. Expected Adapted Perceived Academic Control				
10. Peer Interaction, High School		X	X	
11. Faculty Interaction, High School			X	
12. Faculty Encouragement, High School				
<b>Follow-Up Survey Measures</b>				
13. Self-Efficacy, Social Adjustment			X	
14. Self-Efficacy: Academic Adjustment			X	
15. Mattering: Campus	X	X	X	
16. Sense of Belonging: Campus	X		X	
17. Mattering: Learning Community			X	
18. Validation				
19. Adapted Perceived Academic Control				
20. Mentoring Support				
21. Engagement—First-Year Seminar Courses				
22. Career Decision Making Self-Efficacy				
23. Academic Peer Interaction, Non Learning Community			X	
24. Academic Peer Interaction, Learning Community About Non Learning Community			X	

Scale	Change Response Options	Item(s) Dropped	Item(s) Added	Items(s) Reworded
25. Academic Peer Interaction, Learning Community			X	
26. Social Peer Interaction, Non Learning Community			X	
27. Social Peer Interaction, Learning Community About Non Learning Community			X	
28. Social Peer Interaction, Learning Community			X	
29. Engagement—Shared Academic Courses				
30. Faculty Course-Related Interaction, Non Learning Community			X	
31. Faculty Course-Related Interaction, Learning Community About Non Learning Community			X	
32. Faculty Course-Related Interaction, Learning Community			X	
33. Faculty Noncourse-Related Interaction, Non Learning Community			X	
34. Faculty Noncourse-Related Interaction, Learning Community About Non Learning Community			X	
35. Faculty Noncourse-Related Interaction, Learning Community			X	
36. Faculty Encouragement, Non Learning Community				
37. Faculty Encouragement, Learning Community About Non Learning Community				
38. Faculty Encouragement, Learning Community				
39. Educational Resiliency				X
40. College Knowledge			X	
41. Staff Care and Support				
42. Sense of Belonging: Learning Community	X		X	
43. Academic Self-Efficacy: Study Skills (T1)				
44. Peer Collaboration in Shared Academic Courses (T1)				
45. Academic Validation				
46. Interpersonal Validation				
47. Financial Stress				

Description of Approach: Self-Efficacy: Expected Social Adjustment:

Between administrations, two new items were added to this scale, and two items were reworded. We first checked the functioning of the reworded items to determine whether the modified wording altered the functioning of the items. To do this, we scaled the Cohort 2015 data and Cohort 2016 data (with the six common items only) separately to examine item diagnostics (e.g., item difficulties, hierarchies, fit statistics, point-measure correlations) of items that were reworded between administrations. Our examination of these items' functioning indicated that the items generally appeared to fit and contribute equally to the scale in both versions; however, the relatively small differences in item-difficulty measures across versions were sufficient to



affect the hierarchy of the items between versions. In other words, the rewording of the items was associated with a small shift in the relative difficulty of endorsing items from new to original wording. The analysis team opted for a conservative approach to account for item wording (to not treat those items as common between surveys). Specifically, we scaled the Cohort 2016 data to capture the item parameters from that survey and scaled the full data set (Cohort 2015 and Cohort 2016) anchored with fixed parameters for only the four items that were common to both surveys (present on both versions with exact wording). The other four items were allowed to “float” (i.e., item difficulties were not anchored for the four items that were not common).

We examined diagnostics to confirm that displacement statistics were small, which indicate that the difference between the item difficulties anchored and floating (if no anchoring was applied) is minimal. All item difficulties and thresholds presented in this memo represent the final parameters used for scaling (which may be anchored or unanchored).

#### Description of Approach: Self-Efficacy Expected Academic Adjustment:

Three new items were added to this scale between survey administrations; therefore, students in Cohort 2015 did not receive the additional items in their survey. To account for this difference, we used anchoring techniques to apply fixed item parameters (difficulties and thresholds) from the full set of items to the partial set of items.<sup>14</sup> Specifically, we scaled the data from the Cohort 2016 survey only (the survey with the full set of 15 items), captured the item parameters from that survey, and scaled the full data set (Cohort 2015 and Cohort 2016) with those fixed parameters.

We examined diagnostics to confirm that displacement statistics were small (indicating that the difference between the item difficulties anchored and floating [if no anchoring was applied] is minimal). All item difficulties and thresholds presented in this memo represent the final (anchored) parameters used for scaling.

#### Description of Approach: Expected Matterings: Campus:

Between administrations, changes were made to the items (items were dropped, added, and reworded) and the response options (4-point options were changed to 7-point options) for these scales. Changes to the response options were addressed prior to changes in the items.

*Response Options.* First, we scaled data from the Cohort 2016 survey (7-point response options) and Cohort 2015 survey (4-point response options) separately to descriptively examine the functioning of the 7-point and 4-point response option scales. The pattern of use for the 7-point scale indicates that too many response options were offered to respondents (i.e., that respondents did not necessarily see each as distinct), and that categories could be collapsed as follows to most closely match the pattern of use across versions:

- Values of 1 and 2 recoded as 1
- Values of 3 and 4 recoded as 2
- Values of 5 and 6 recoded as 3

---

<sup>14</sup> Anchoring is a technique where known item parameters are fixed in the scaling process, allowing scales with common items (in this case, the items that appeared in all versions of the survey) to be equated.

- Value of 7 recoded as 4

We collapsed data from the Cohort 2016 survey as indicated and rescaled the 2016 data file with the collapsed 4-point response options.

*Item Changes.* We rescaled the Cohort 2016 survey (collapsed 4-point response options) with only the six items common to both cohorts (i.e., dropping items that were introduced to the survey between administrations) to examine item diagnostics (e.g., item difficulties, hierarchies, fit statistics, point-measure correlations) of items that were reworded between administrations. Our examination of these items' functioning indicated that the rewording was sufficiently minor and did not affect the items' position in the scale. Item rewording was disregarded from this point forward (items were treated as common).

To account for the addition of items between administrations, we rescaled the Cohort 2016 data (full item set, collapsed 4-point response options) to capture the item parameters from that survey, and scaled the full data set (Cohort 2015 and Cohort 2016) anchored with those fixed parameters.

We examined diagnostics to confirm that displacement statistics were small (indicating that the difference between the item difficulties anchored and floating [if no anchoring was applied] is minimal). All item difficulties and thresholds presented in this memo represent the final (anchored) parameters used for scaling.

#### Description of Approach: Expected Sense of Belonging:

Between administrations, items were added to the survey, and the response option scale was changed (from 4-point options to 7-point options). Changes to the response options were addressed prior to changes in the items.

*Response Options.* First, we scaled data from the Cohort 2016 survey (7-point response options) and Cohort 2015 survey (4-point response options) separately to descriptively examine the functioning of the 7-point and 4-point response option scales. The pattern of use for the 7-point scale indicates that too many response options were offered to respondents (i.e., that respondents did not necessarily see each as distinct), and that categories could be collapsed as follows to most closely match the pattern of use across versions:

- Values of 1 and 2 recoded as 1
- Values of 3 and 4 recoded as 2
- Values of 5 and 6 recoded as 3
- Value of 7 recoded as 4

We collapsed data from the Cohort 2016 survey as indicated and rescaled the 2016 data file with the collapsed 4-point response options.

*Item Additions.* To account for the addition of items between administrations, we captured the item parameters from the Cohort 2016 survey and scaled the full data set (Cohort 2015 and Cohort 2016) anchored with those fixed parameters.

We examined diagnostics to confirm that displacement statistics were small (indicating that the difference between the item difficulties anchored and floating [if no anchoring was applied] is minimal). All item difficulties and thresholds presented in this memo represent the final (anchored) parameters used for scaling.

#### Description of Approach: Expected Mattering Learning Community:

This scale was only presented in the Cohort 2016 baseline survey; therefore, no changes were made over time. No data management or anchoring was needed for this scale.

#### Description of Approach, Mattering: Campus and Sense of Belonging

Over time, changes were made to the items (dropped and added) and the response options (4-point options to 7-point options) for these scales. Changes to the response options were addressed prior to changes in the items.

First, we scaled all timepoints after 2016 (7-point response options) to descriptively examine the functioning of the 7-point response options. For both scales the pattern of use indicates that too many response options were offered to respondents (i.e., that respondents did not necessarily see each as distinct)—consistent with the findings and recommendations from the 2018 psychometric memo. We then scaled the 2016 data file (4-point response options) independently to examine how the four categories were used by respondents, and determined that (for both scales), based on the distribution of use of 7-point response options and 4-point response options, data from the later administrations should be collapsed as follows for consistency:

- Values of 1 or 2 recoded as 1
- Values of 3 or 4 recoded as 2
- Values of 5 or 6 recoded as 3
- Values of 7 recoded as 4

After collapsing data with 7 points and reexamining diagnostics, we followed the same anchoring methods detailed above to create directly comparable scale scores across all timepoints and cohorts.

#### Description of Approach: Scales With Items Added Over Time

New items were added to each of these scales, as indicated in the table above, between the 2016 and 2017 survey administrations; therefore, students in Cohort 2015 did not receive the additional items until timepoint 2. To account for this difference, we used anchoring techniques to anchor item difficulties and thresholds. Anchoring is a technique where known item parameters are fixed in the scaling process, allowing scales with common items (in this case, the items that appeared in all versions of the survey) to be equated.

To equate these scales across time, we first calculated item and scale parameters (item difficulties and rating scale thresholds) based on all surveys after 2016, then ran the scaling for the full data file (all cohorts and all timepoints) using those anchored values for item difficulties

and thresholds.<sup>15</sup> We examined diagnostics to confirm that displacement statistics were small (indicating that the difference between the item difficulties anchored and floating [if no anchoring were applied] is minimal). All item difficulties and thresholds presented in this memo represent the final (anchored) parameters used for scaling.

#### Description of Approach: Scales That Did Not Change Over Time

For some scales, as indicated in the table above, no changes were made over time; therefore, the approach used to calculate directly comparable scores was straightforward: scaling was run on the full data set (all cohorts and timepoints). No additional data management or anchoring was needed for these scales.

#### **Notes About Measurement Error**

Person ability measures calculated from the Rasch analysis presented here are estimates of a person's score on the latent trait, based on their responses to survey items. These person ability measures (in logits) are provided in the accompanying data file, along with their associated standard errors. In general, we know the least about persons at the tails of the ability distribution, so we expect standard errors to be largest at the tails. It is worth noting that the items added to the TSLC baseline survey between administrations were sometimes at the tails of the item difficulty distributions (as listed above), providing more information about persons at the tails of the distribution in cohort 2016 (i.e., we expect that precision is increased for cohort 2016 extreme scores). We encourage data users to consider standard errors when using these person ability measures in analyses.

---

<sup>15</sup> A variety of anchoring approaches can be used to calculate directly comparable scores across survey administrations. Approaches are similar and tend to result in very little difference in scores. To confirm consistency between methods, we applied an alternative approach where we first ran the scaling separately for the different versions (before and after changes), and examined diagnostics to confirm consistency across versions (Step 1); calculated rating scale thresholds using the full data file (all cohorts and timepoints) (Step 2); applied those anchored thresholds to the first version of the survey (Cohort 2015, timepoint 1), and calculated item difficulties (Step 3); and scaled the full data set with the anchored rating scale thresholds from Step 2 and the anchored item difficulties from Step 3 (Step 4). Differences across these two methods were very minimal, with an average difference score of 0.02 and a range of -0.04–0.16 logits.

## References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573.
- Granger, C. (2008). Rasch analysis is important to understand and use for measurement. *Rasch Measurement Transactions*, 21(3), 1122–1123.
- Linacre, J. M. (2018). *WINSTEPS: Rasch Measurement Computer Program* [Computer software]. Chicago, IL: Winsteps.com.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.