

University of Colorado End of First Year Survey

Part of an application for the role of data analyst with SA-PANDA

Jayson Webb, PhD

2023-09-24

Introduction

This report is in response to a request to analyze the End of First Year Survey administered by the Student Affairs Planning, Assessment & Data Analytics (SA-PANDA) office at the University of Colorado. This is in partial fulfillment of an application for the role of data analyst with SA-PANDA. The analyses will focus on how ratings of belonging and mattering survey items differ within one student group and one demographic identity. The code used to generate this report can be downloaded from this [github repository](#).

Process

This section describes my approach to this analysis in terms of cleaning and organizing the data, software and statistical techniques used and the student group and demographic identity selected for the analysis.

Cleaning and Organizing the Data

The four biggest issues in cleaning and organizing the data were 1. Deciding on the approach to handling missing values; 2. Transforming verbal item ratings into numerical values (e.g. Strongly agree = 6); 3. Shortening column names for survey items so they looked better on graphs and were easier to use as variable names in statistical analyses and 4. Putting the data into a [tidy format](#) for some analyses, meaning question type (e.g. belonging) became a column heading, the specific question became the value (e.g. “Fit-in”) and the rating occupied a separate column. The latter three issues were straightforward, but I will say a bit more about analyzing the ratings as numerical data at the end of this section.

Missing data

The website [MeasuringU](#) provides an excellent summary of missing data considerations for survey analysis. For data summaries, t-tests and correlations, I took the ‘pair-wise deletion’ approach. That means that, for example, a slightly different set of people may have provided answers for the survey items “I belong at CU Boulder” and “I fit in well at CU Boulder”. For each question, I analyzed the responses for the people that did provide answers, resulting in a slightly different sample size for each question. The four ANOVAs I performed required a ‘list-wise deletion’ approach, which means that for a given analysis which included both a subset of questions (e.g. the belonging questions) and a subset of participants (e.g. those who provided answers to the mentor question), only respondents who provided answers to all questions were included. Luckily, for a given analysis, there wasn’t a lot of variability from question to question in terms of who provided answers.

Perhaps the most important question about missing data is whether missing data in one variable (e.g. “I belong at CU Boulder”) is systematically related to missing data in another variable (e.g. “Do you have someone at CU Boulder you consider a mentor?”). For example, if people who respond “No” to the mentor question are less likely to answer the belonging question than those who respond “Yes”, then there is a systematic effect to the pattern of missing data that could affect the conclusions. I applied the following logic to assess if missing data is systematic and if I think it would affect the conclusions:

- Does one of the variable levels (e.g. responded “No” to the mentor question) have a significantly different response rate than another level (e.g. “Yes” responders to that same question)?
- If one group is less likely to respond, would the expected effect be in the opposite direction to the observed effect? For example, I would expect “No” responders to the mentor question to generally feel a lower sense of belonging, which is in the same direction as the observed effect, so it would not change the conclusion.
- Is the size of the observed effect such that a non-response effect would likely change it?

In the Results section, I won’t mention the missing data issue unless I think it could be a problem.

Coding survey responses as numerical

For this report, I analyzed the average numerical value of belonging and mattering questions, coding the responses as 1 through 6, rather than something like percent top-box or percent top-two box responses (e.g. percent who chose agree or strongly agree). In my career I have used both methods, depending on the preferences and practices of the organization I worked for and the purpose of the analysis. Because there is an emphasis on looking for statistically significant differences in this analysis, I chose to analyze the numerical values via ANOVA and t-test because these approaches typically have more statistical power than analyzing top-box or top-two box percentages. One can find controversy online about whether survey responses

should be treated as ordinal or interval data requiring non-parametric statistical methods. I have also seen ample convincing arguments that it is proper to take averages and use parametric methods with responses from Likert-type items. A full literature review on this issue is beyond the scope of this paper, but the arguments all rely on the central limit theorem and the fact that the sampling distribution of means converges to a normal distribution as the sample size increases, regardless of the nature of the underlying distribution.

Software and statistical techniques

Python was used for exploratory data analysis and to generate the graphs and statistical analyses in the report. As mentioned in the introduction, the code used to generate the report is in a “.qmd” file in the [github repository](#) for this analysis. The .qmd file is a [Quarto](#) file. I run Quarto inside of RStudio to both write the narrative for the analyses as well as perform the analyses with Python code. This follows the [literate programming](#) philosophy where documentation & code are combined in a resource that can be examined in order to improve transparency and reproducibility of analyses. It also improves the flexibility and agility of deliverables. With small tweaks to the metadata in the file, the same code and text can be delivered as a PDF, an [interactive html document](#), a slide deck or other formats.

The primary statistical analyses used were 1. mixed two-way ANOVA as an omnibus F test prior to planned comparisons and 2. t-tests to conduct planned comparisons following a significant omnibus F test.

The mixed two-way ANOVAs had one repeated measures factor (belonging or mattering, with the different questions as levels of the repeated factor) and one between-participants factor (mentor or entry type). If the between-participant factor or the interaction of the repeated measures factor with the between-participants factor was statistically significant beyond the $p = 0.05$ level, planned comparisons were performed. When the interaction was significant, I only report that and not the significance of the individual factors, following convention.

All post-hoc comparisons were considered planned, which meant they were decided upon prior to any tests being conducted, conditional on a significant omnibus F test and were driven by the theoretical question at hand. Because they were planned comparisons, no family-wise error correction was done.

Student Group and Demographic Identity Used in the Analyses

I chose to look at mentoring (student group) and entry type (demographic). For mentoring, I chose to focus planned comparisons on those who responded “Yes” versus “No” to the question “Do you have someone at CU Boulder you consider a mentor?”. For entry type, I’m comparing transfer students to first-time college students - the only two options for that variable. The criteria for choosing these were:

- **Actionable** - Does it seem like there are clear implications for action for differences in belonging and mattering? In the case of having a mentor, the action seems clear: having a mentor makes a big difference in belonging and mattering, so get more people to have mentors. For entry type, you can't change someone's status as a transfer student, but you can make an effort to focus on those students.
- **Understandable** - Are the differences in student groups easy to understand or is more work needed? To me, the mentoring and entry type characteristics seemed straightforward. I was strongly considering race/ethnicity, but one thing that makes that variable challenging is the ability to select multiple identities. More study is needed before I would be comfortable combining or excluding certain levels of that variable.
- **Lots of support in the data** - In this data, the smaller categories (no mentor, transfer student) had over 150 respondents for any one belonging or mattering question. For ethnicity, some of the potential categories of interest had fewer than 25 responses for some questions.

Results

Per the assignment, this section provides a descriptive analysis of the belonging and mattering survey responses overall. This will provide a sense of the baseline levels for average ratings prior to examining the effects of mentor status and entry type. Figures 1 and 2 show the average ratings for the responses to the four belonging and three mattering questions.

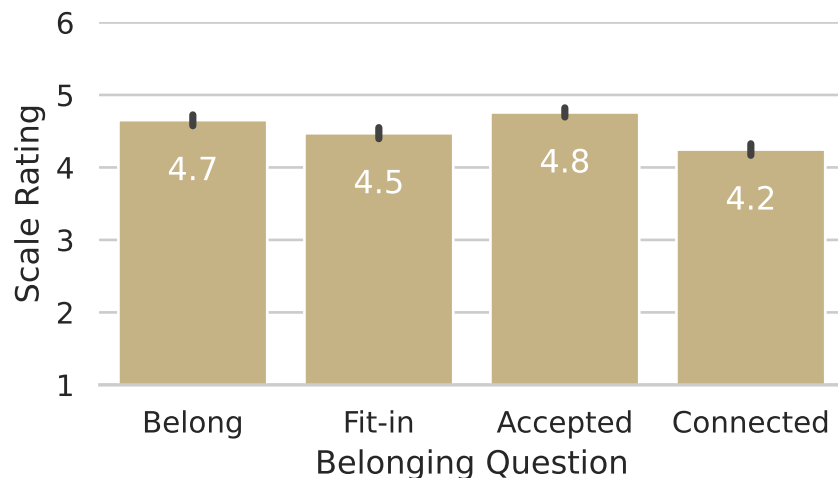


Figure 1: Overall Belonging Ratings - Belong: I belong at CU Boulder, N=1152; Fit-in: I fit in well at CU Boulder, N=1174; Accepted: People at CU Boulder accept me, N=1169; Connected: I feel a connection with the CU Boulder community, N=1179

Error bars indicate the 95% confidence interval for the average rating. Respondents rated agreement using the following response choices: 1=Strongly disagree, 2=Disagree, 3=Slightly disagree, 4=Slightly agree, 5=Agree, 6=Strongly agree.

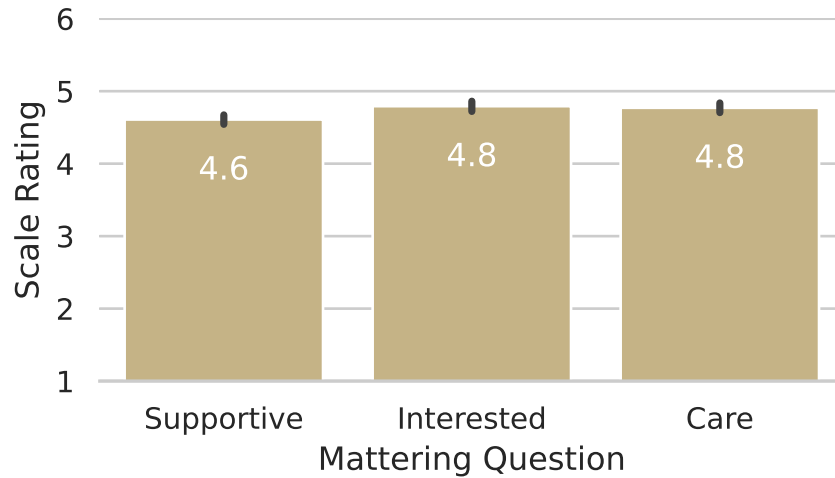


Figure 2: Overall Mattering Ratings - Supportive: People on campus are generally supportive of my individual needs, N=1161; Interested: There are people on campus who are genuinely interested in me as a person, N=1179; Care: There are people on campus who care about my future, N=1169

The next sections will look at differences in feelings of belonging and mattering within two different student classifications. First, I'll examine the effect of having a mentor vs not having a mentor and then address the effect of being a transfer student.

Having a mentor improves sense of belonging and mattering

Figure 3 shows that belonging ratings differ by a student's response to the question *Do you have someone at CU Boulder you consider a mentor?*, with ratings for those who responded "No" (54% overall) being lower than for "Yes" (33% overall) or "Not sure" (13% overall) respondents.

A mixed two-way ANOVA shows that the pattern of differences between mentor question responses ("Yes", "No", "Not sure") varies by belonging question (Belong, Fit-in, Accepted, Connected). This is indicated by a significant interaction between belonging question and mentor response, $F(6, 3360) = 6.45$, $p = 8.4 \times 10^{-7}$. Planned comparisons focused on the difference between the "Yes" and "No" responses to the mentor question and show that average ratings for "No" are significantly lower than for "Yes" responses for each belonging question. This can be seen visually in Figure 3 by the non-overlapping of the 95% confidence intervals for "Yes" and "No" responses. The results of the planned comparisons are shown in Table 1.

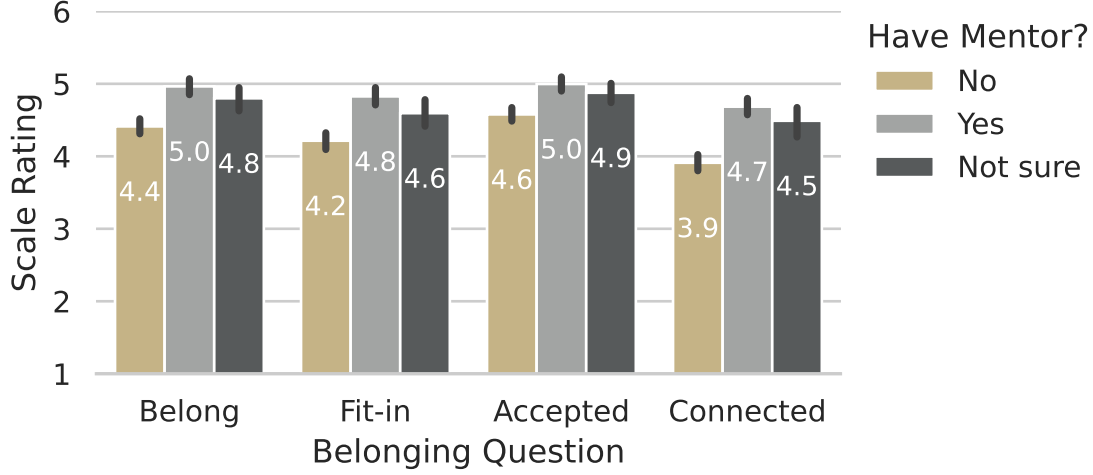


Figure 3: Effect of having a mentor on belonging ratings. Belong (No:N=606, Yes:N=396, Not sure:N=150) | Fit-in (No:N=623, Yes:N=398, Not sure:N=153) | Accepted (No:N=621, Yes:N=394, Not sure:N=154) | Connected (No:N=629, Yes:N=397, Not sure:N=153)

Table 1: Results of planned comparisons of “Yes” vs “No” responses to the mentor question for each belonging question

Belonging Question	Difference (Yes-No)	Cohen’s d	df	t-value	p-value
Belong	.55	.43	1000	6.7	$p < .0001$
Fit-in	.61	.46	1019	7.2	$p < .0001$
Accepted	.42	.38	1013	5.8	$p < .0001$
Connected	.77	.57	1024	8.9	$p < .0001$

The difference column shows that the largest difference in belonging ratings between “Yes” and “No” respondents was for the “Connected” question, with a difference in mean ratings of .77. [Cohen’s d](#) is a standardized difference based on the pooled standard deviation used in the t-test and is a common measure of [effect size](#) in the social sciences.

Figure 4 shows that mattering ratings also differ by responses to the mentor question. The interaction between mentor question and mattering questions was not significant at the $p < .05$ level, $F(4, 2266) = 2.01$, $p = .091$, indicating that the pattern of differences among the mentor responses did not differ by mattering question. But, the effect of the mentor question was significant, indicating that there were differences in average question rating based on response to the mentor question, $F(2, 1133) = 38.23$, $p = 8.6 \times 10^{-17}$. Table 2 shows the results of planned comparisons.

The “Yes” vs “No” comparisons shown in Table 2 were significant for all mattering questions, as was the case for the belonging questions.

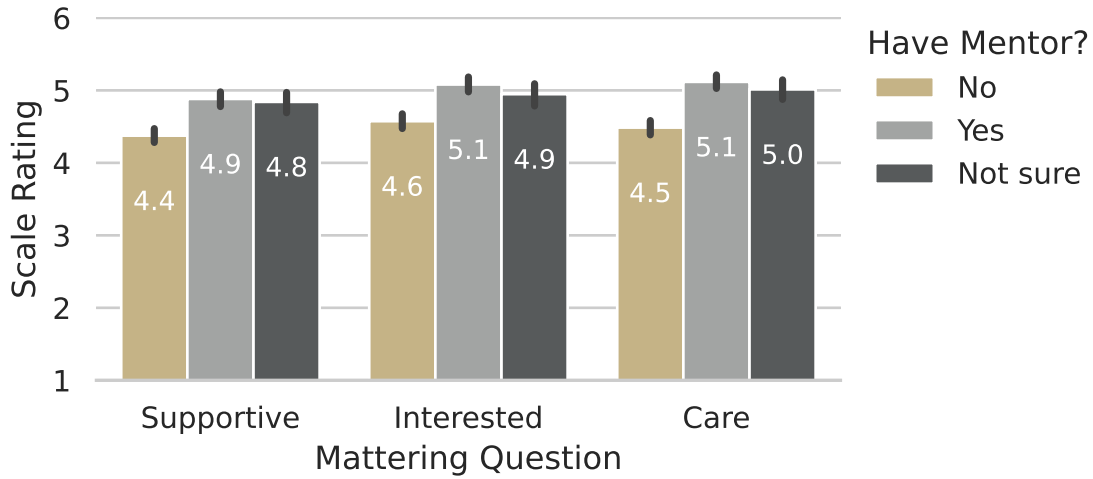


Figure 4: Effect of having a mentor on mattering ratings. Supportive (No:N=613, Yes:N=397, Not sure:N=151) | Interested (No:N=630, Yes:N=397, Not sure:N=152) | Care (No:N=617, Yes:N=401, Not sure:N=151)

Table 2: Results of planned comparisons of “Yes” vs “No” responses to the mentor question for each mattering question

Mattering Question	Difference (Yes-No)	Cohen’s d	df	t-value	p-value
Supportive	.51	.44	1008	6.9	$p < .0001$
Interested	.51	.43	1025	6.7	$p < .0001$
Care	.63	.56	1016	8.7	$p < .0001$

Having a mentor seems to clearly improve a student’s sense of belonging and mattering.

Transfer students have a lower sense of belonging and mattering than first-time college students

Figure 5 shows the effect of being a transfer student (about 17% of respondents) on belonging ratings. As with the mentoring question, there was a significant interaction between belonging question and transfer status, $F(3, 3363) = 5.98$, $p = 4.7 \times 10^{-4}$.

Planned comparisons, shown in Table 3, indicate that the effect of transfer status applies to every question, although the p value for the Belong question was greater than 0.05, but not by

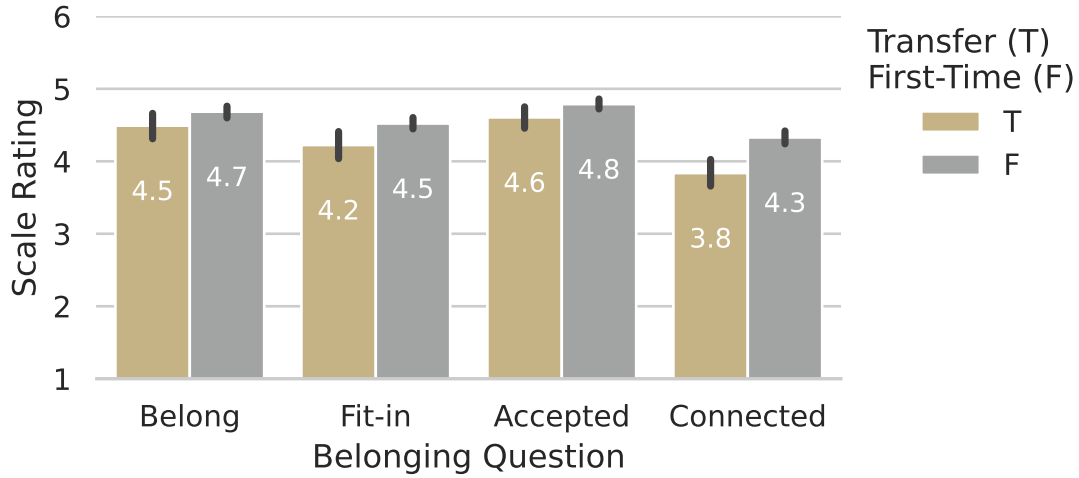


Figure 5: Effect of being a transfer student on belonging ratings. Belong (Transfer:191, First-time:961) | Fit-in (Transfer:196, First-time:978) | Accepted (Transfer:193, First-time:976) | Connected (Transfer:197, First-time:982)

much. Compared to the effect of mentor status, the differences and standardized differences (Cohen's *d*) are smaller. For the mentor question, the differences were all greater than .4, whereas for transfer status, only the Connected question shows an effect of that size (.49).

Table 3: Results of planned comparisons between transfer and first-time college student responses for each belonging question

Belonging Question	Difference (Yes-No)	Cohen's <i>d</i>	df	t-value	p-value
Belong	.19	.15	1150	1.9	.056
Fit-in	.30	.22	1172	2.8	.0046
Accepted	.18	.17	1167	2.1	.035
Connected	.49	.35	1177	4.5	$p < .0001$

Figure 6 shows the effect of being a transfer student on mattering ratings. As with the mentoring question, the interaction between transfer status and mattering questions was not significant but the effect of transfer status was, $F(1, 1134) = 13.62$, $p = 2.4 \times 10^{-4}$. Table 4 shows the results of planned comparisons between transfer and first-time college student responses for each mattering question. All of the comparisons shown in Table 4 are statistically significant. As with the belonging questions, the effect of being a transfer student wasn't as large as the effect of not having a mentor.

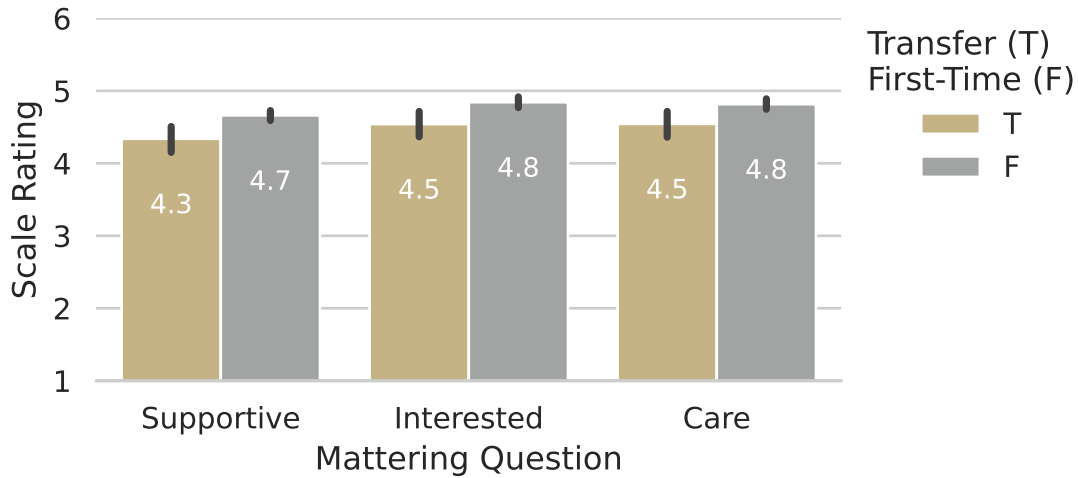


Figure 6: Effect of being a transfer student on mattering ratings. Supportive (Transfer:189, First-time:972) | Interested (Transfer:194, First-time:985) | Care (Transfer:193, First-time:976)

Table 4: Results of planned comparisons between transfer and first-time college student responses for each mattering question

Mattering Question	Difference (Yes-No)	Cohen's d	df	t-value	p-value
Supportive	.32	.28	1159	3.6	.0004
Interested	.30	.26	1177	3.3	.0011
Care	.27	.24	1167	3.0	.0026

Summary and Implications

This analysis identified two group differences that affect student belonging and mattering ratings. Not having a mentor (54% of respondents), compared to having a mentor, resulted in lower belonging and mattering ratings. Also, being a transfer student (17% of respondents) resulted in lower belonging and mattering ratings. The effect of not having a mentor was larger, and affects a larger percentage of the student body, than the effect of being a transfer student, but the effect was statistically significant for both and affected all questions related to belonging and mattering.

These results have practical implications in the context of the college campus environment. In the case of having a mentor, the action seems clear: get more students to have mentors. In terms of being a transfer student, make an effort to focus on those students, including getting them a mentor.