Measuring Alumni Career Outcomes
A Validity Study

Mason R. Jones, M.A.
James Madison University
Doctoral Assistant,
Center for Assessment & Research Studies (CARS)
jones5mr@jmu.edu | ORCiD

Megan R. Good, Ph.D.
James Madison University
Executive Director, CARS
Assistant Professor, Graduate Psychology
LinkedIn | ORCiD

Sara J. Finney, Ph.D.
James Madison University
Associate Director, CARS
Professor, Graduate Psychology
LinkedIn | ORCiD

Abstract: Colleges and universities use surveys like the National Alumni Career Mobility (NACM) to evaluate how well they prepare graduates for careers. This study tested two NACM scoring models—a five-factor model and a bifactor model—and evaluated whether results were consistent across first-generation and non-first-generation alumni. Findings supported the simpler five-factor model (i.e., support for computing five subscale scores), suggested removing a problematic item, and showed that the survey functioned consistently across groups. These findings provide the first peer-reviewed guidance on NACM scoring and support the use of subscale comparisons to inform student career development efforts.

Keywords: career mobility, career outcomes, first-generation alumni, confirmatory factor analysis, measurement invariance

---

To address concerns about the value of higher education, it is crucial to provide college students and stakeholders with career outcome data that shows how well institutions prepare graduates for professional success. Students anticipate that their college education will lead to successful careers (Blaich & Wise, 2021). However, studies that connect career data to college experiences are scarce due to two primary research challenges. First, defining a positive career outcome is complex. One frequently studied career outcome is salary, an important but limited facet of career success. Other dimensions of career outcomes include career satisfaction, economic mobility, and career mobility (i.e., the ability of an individual to move between or within different jobs or roles; Büchel & Mertens, 2004; Dumford & Miller, 2017; Miller, 1984; Sicherman & Galor, 1990). Thus, institutions need to measure multiple aspects of career outcomes when studying post-graduate success.

A second obstacle when studying post-graduate career outcomes is logistical. Higher education institutions often lack up-to-date contact information for their alumni. This challenge has been addressed by third-party entities that obtain institutional alumni career data for a fee (e.g., Measuring Higher Education Outcomes–Gallup-Purdue Index, n.d.). One such group, The Career Leadership Collective (CLC), developed a survey to measure career outcomes and relate them to college experiences. The National Alumni Career Mobility (NACM) survey defines career mobility and helps institutions articulate the value of their degrees (CLC, 2022). Since this study was conducted, the NACM survey was acquired by Lightcast.

National Alumni Career Mobility Survey (NACM)

The NACM survey is a proprietary tool developed in 2018 (CLC, 2022). NACM was designed to help higher education administrators "adapt their career development practices and to equitably guide students toward more fulfilling careers and lives" (CLC, 2022, p. 4). NACM has been administered to alumni from over 50 public and private U.S. institutions. The Collective has published national public reports, presented NACM data in webinars, and provided data visualization services to institutions. As part of the overall survey, 25 items are typically scored using five subscales: Career Pathway Preparation (CPP), Career Satisfaction (CS), Community Engagement (CE), Economic Mobility (EM), and Institutional Career Investment (ICI; see Appendix).

The CPP subscale measures alumni's readiness for their chosen career after postsecondary education. The feeling of contentment that alumni experience is measured by CS. CE gauges the extent to which alumni actively participate in their local communities, whereas EM compares current income or earning potential with the economic conditions of the households in which they grew up. Lastly, ICI reflects the perceived level of support that alumni received from their institutions in terms of career development. NACM reports consistently present these five subscale scores in campus deliverables and national dataset reporting. In addition, a career mobility index (CMI) score was reported, encompassing the CPP, CS, and EM subscales. Also, NACM scores were routinely disaggregated by demographic variables (e.g., first-generation status) to draw attention to potential equity issues.

This study examined the validity of the NACM factor structures. Validity ensures that the way we interpret scores from the survey is supported by research and evidence (American Educational Research Association et al., 2014). Benson (1998) proposed a three-step process for evaluating validity. The first step, the substantive stage, focuses on developing survey items based on a clear theoretical framework and empirical research. The second step, the structural stage, evaluates the internal structure of the survey, often through statistical methods like factor analysis, to ensure that items measure the intended constructs. The third step, the external stage, involves testing a priori stated hypotheses about the relation between the scores produced by the measure and other constructs (e.g., expected relationships). We used this process in the present study of NACM.

Evaluation of Prior Validity Information

There are no published studies on the NACM survey, despite its widespread use. Analysis of the national reports and consultations with The Collective indicates that the measure lacks a strong theoretical foundation (CLC, 2022). Typically, a validity study would end here. However, the instrument is administered to thousands of alumni every year. The use of this instrument will not cease, so it is crucial that scores reported to universities and stakeholders are reliable and trustworthy. Previous internal psychometric reports of the NACM used structural equation modeling (SEM) to explore how the survey should be scored (CLC, 2021a, 2021b). Specifically, two competing models were tested: a five-factor model, which assumes the survey measures five distinct areas, and a bifactor model, which includes the CMI (see Figure 1). However, previous reports lacked details on which model fits the data better, leaving the scoring process unresolved.

When evaluating a measure, it is important to provide clear and detailed information so that others can understand and replicate the analysis. However, the methodology reported by The Collective leaves several gaps. For example, it is unclear whether the five-factor model was tested using exploratory or confirmatory methods, or which statistical tools and techniques were used to analyze the data. Key details such as how the data were screened and how reliability was assessed were excluded from the methodology. This limits our ability to understand or replicate their findings. Details provided indicated that Cronbach's alpha was used to assess reliability in this study. While alpha is a common choice, omega is often considered more appropriate in SEM.

Additionally, the reports failed to include measures of local fit, such as correlation residuals, which are critical for evaluating SEM model fit. Finally, the raw data was not shared, making it difficult to evaluate their results independently. As a result, further analysis is needed to compare the five-factor and bifactor models to determine which best fits the data, as this decision directly impacts how scores are generated and interpreted. Identifying the appropriate model ensures that NACM subscale scores accurately reflect the intended constructs and can be meaningfully used in practice.

Measurement Invariance

In national reports, group comparisons are made using the five NACM subscales and the CMI to explore equity concerns. Specifically, data is broken down by demographic variables, such as first-generation status. For these comparisons to be meaningful, the survey must produce comparable scores across different groups. This concept, called measurement invariance, involves confirming that any score differences observed between groups truly reflect real differences between groups, rather than issues with the survey itself (Kline, 2023).

Measurement invariance is tested in three steps. First, configural invariance checks whether the survey questions group together similarly for each demographic. For example, the five-factor model should apply equally well to first-generation and non-first-generation students. If this holds, the next step is to test metric invariance, which evaluates whether the survey items measure the underlying concept to the same degree for each group. In other words, does each item represent the associated construct (e.g., career satisfaction) equally well for first-generation and non-first-generation students? If metric invariance is not supported, different groups may interpret some items differently, raising validity concerns.

If configural and metric invariance standards are met, a researcher can test the most stringent test of invariance–scalar. Scalar invariance tests whether students with the same level of the underlying concept (e.g., career satisfaction) answer survey items similarly, regardless of their group. When both metric and scalar invariance are supported, we can confidently compare average scores between groups (e.g., comparing career satisfaction across first-generation and non-first-generation students). For example, if scalar invariance holds, first-generation and non-first-generation alumni with the same level of career satisfaction should report similar scores on the CS items. If it does not hold, one group might systematically rate their satisfaction higher or lower, even if their actual experiences are the same. Without evidence of measurement invariance, differences observed between groups may not reflect true differences and could instead result from inconsistencies in how the survey functions across groups.

First-Generation College Students

We selected first-generation status as the demographic variable for invariance testing because it is a key comparison group in national reports for NACM and is supported by extensive literature documenting distinct differences in college experiences between first-generation and non-first-generation students. First-generation college students have traditionally been defined as those with neither parent having obtained a postsecondary degree (Choy, 2001; Gable, 2021). First-generation and non-first-generation students differ in their success, as measured by graduation rates, retention rates, college experiences, and first-destination career outcomes (i.e., landing a first job out of college or being accepted into graduate school). First-generation students, on average, tend to be lower on these metrics (Chien et al., 2016; Kuh et al., 2006; Manzoni & Streib, 2019; Pascarella et al., 2004).

Given the unique challenges faced by first-generation students in college, conceptual differences in NACM subscale scores may emerge based on first-generation status. Specifically, conceptual differences may emerge in CPP and ICI because first-generation students often navigate unique college experiences, including opportunities to bridge cultural contexts and develop resilience in adapting to new environments (Gable, 2021). In addition, studies on short-term career outcomes for first-generation students yield mixed results, indicating the possibility of different short-term outcome experiences and conceptualization of items on the EM subscale (Chien et al., 2016; Choy, 2001). In particular, items referencing earning potential relative to the household in which respondents grew up may function differently across the two groups (different factor loadings or intercepts) if these items are conceptualized differently. In sum, if a model is championed, we would not be surprised by a lack of measurement invariance for the CPP, ICI, and EM subscales.

The Current Study

Given the prevalence of the NACM and the dearth of studies examining its psychometric properties, we had two goals. First, we evaluated two proposed but insufficiently studied factor structures to help guide the appropriate scoring of the NACM: a correlated five-factor structure and a bifactor structure that supports the construction of the CMI score. Second, we tested measurement invariance across first-generation and non-first-generation students. This is the first peer-reviewed published study of the NACM.

Method

Data Collection Procedure

The Collective conducted data collection for NACM in three phases. First, an online survey portal was created for each higher education institution (*N* = 77). Institutions sent messages to alumni based on their graduation year, directing them to the survey portal. Second, a social media data scrape was performed on all alumni who did not complete the survey when prompted by their institution. Third, alumni five and ten years out of college were emailed requesting survey engagement.

Participants

Data were collected from alumni from 77 different institutions across the United States in 2021 and 2022. The full dataset had 25,763 respondents. Only undergraduate alumni respondents from the 2021 (2016 and 2011 graduates) and 2022 (2017 and 2012 graduates) cohorts were retained, resulting in a sample size of 19,819. Respondents who did not have full responses on all 25 NACM items were removed, yielding a sample size of 19,313. Most records deleted at this step were missing information on most items. Respondents were also excluded if they had missing data on the “first-generation status” variable or indicated that they did not want to answer the “first-generation status” question, resulting in a final sample of 16,481. Data from the 2021 cohort included alumni who graduated in 2016 (*N* = 5,443) and 2011 (*N* = 4,298). Data from the 2022 cohort included alumni who graduated in 2017 (*N* = 5,456) and 2012 (*N* = 4,116). We split the dataset (*N* = 16,481) in half, generating two random stratified samples based on first-generation status. Having two independent samples allowed us to evaluate if any item-data misfit replicated. We were uncomfortable evaluating the quality of NACM items without replication in an independent sample.

Samples 1 and 2 contained 8,241 and 8,240 alumni, respectively. Overall, the majority of both samples were White and self-identified as female. When examining samples 1 and 2 by first-generation status (*N* = 2,537 for both samples), there were more self-identified women among first-generation students in both samples. First-generation students were more diverse than non-first-generation students across samples in terms of racial and ethnic background. These differences were expected, given prior research (Choy, 2001; Hamilton, 2023; Kuh et al., 2006).

Data Analysis

Confirmatory Factor Analysis (CFA)

To evaluate how to score the NACM survey, we tested two models using confirmatory factor analysis (CFA): a five-factor model and a bifactor model. The five-factor model (see Figure 1) assumes the survey measures five distinct areas: Career Pathway Preparation (CPP), Career Satisfaction (CS), Economic Mobility (EM), Community Engagement (CE), and Institutional Career Investment (ICI). In contrast, the bifactor model introduces an additional general factor, the Career Mobility Index (CMI), which accounts for the overlap among CPP, CS, and EM. Since the five-factor model is nested within the more complex bifactor model, we used a chi-square difference test ($\Delta\chi^2$) to determine whether the added complexity of the bifactor model significantly improved fit (Bandalos & Finney, 2019) and thus justified the creation of the CMI. We also examined global fit indices, including the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), and the comparative fit index (CFI; Hu & Bentler, 1998, 1999). These indices provided an overall sense of how well the models aligned with the data and whether either scoring approach (five subscales versus five subscales and the CMI) was appropriate. Additionally, we assessed local fit by examining correlation residuals to identify areas where the model struggled to reproduce relationships between specific survey items, flagging issues when residuals exceeded |.15| (Kline, 2023).[1]

Measurement Invariance

If one of the models (five-factor or bifactor) fits the data adequately, we could test whether it functioned consistently across first-generation and non-first-generation students, a process called measurement invariance. Measurement invariance was evaluated in three stages. First, configural invariance checked whether the overall structure of the model was the same for both groups. Next, metric invariance tested whether the relationships between the survey items and the underlying factors (i.e., factor loadings) were of similar magnitude across groups.[2] Finally, scalar invariance assessed whether students with the same level of a factor, such as CS, would score similarly on the survey regardless of their group. Each stage involved progressively adding constraints to the model and comparing the fit using statistical tests.[3] If strong evidence of measurement invariance was found (e.g., scalar invariance), effect sizes could be calculated representing the difference in the construct across the two generation status groups.[4]

Results

Before delving into the statistical details, we begin with a summary of our key findings: The five-factor model of alumni outcomes was supported for both first-generation and non-first-generation alumni and was preferred over the more complex bifactor model due to its theoretical clarity and parsimony. Across two samples, item 31 ("My salary is enough to pay my bills every month") consistently caused issues and contributed little to measuring CS. Removing this item improved and addressed local misfit issues. The survey functioned equivalently across groups, enabling meaningful comparisons. Latent mean modeling showed that first-generation alumni scored significantly higher on EM (0.72 standard deviations) compared to non-first-generation alumni, with no other practical or significant differences on the other constructs. What follows are the statistical details supporting these findings.

---

[1] Correlation residuals represent the difference between observed relationships and those predicted by the model. Residuals greater than |.15| suggest areas of local misfit that may need attention (e.g., overrepresented or underrepresented relationships; Bandalos & Finney, 2019).

[2] To identify the model, one factor loading per group was fixed to equality, and the variance of each factor was constrained to 1.0 for the first-generation group, which served as the reference group (Widaman & Olivera-Aguilar, 2023).

[3] The chi-square difference test ($\Delta\chi^2$) compares nested models to determine if adding complexity improves model fit significantly. Global fit indices help evaluate overall model fit: SRMR measures average differences between observed and predicted correlations, RMSEA evaluates approximate fit, and CFI compares the target model to a baseline model. Local fit was also examined by reviewing mean residuals, with differences larger than |.25| flagged for further consideration.

[4] Mean residuals reflect differences between observed and predicted item means; deviations larger than |.25| suggest areas where the model may not perform equally across groups.

Data Screening for Samples

We screened the data to determine the best estimation method for our models (Finney et al., 2016). First, we confirmed that the full response scale (1 to 5) was used for all 25 items in both samples, allowing the data to be treated as continuous. Next, we checked for univariate and multivariate normality. None of the 25 items showed significant univariate skewness or kurtosis (see Table 1), but multivariate non-normality was present, as indicated by Mardia's kurtosis values (e.g., 118.73 to 130.20 for samples 1 and 2, respectively).[5] To address this, we estimated models using both unadjusted maximum likelihood (ML) and Satorra-Bentler adjustments (Satorra & Bentler, 1994).[6] Since results were nearly identical across methods, we reported unadjusted ML results for simplicity. All analyses were conducted in R 4.3.1 using the lavaan package (Rosseel, 2012).

Sample 1

Configural Invariance: Evaluating the Five-Factor Model and Bifactor Model

We tested the five-factor and bifactor models to determine the best-fitting structure and, hence, the appropriate scoring of the NACM. The five-factor model provided a strong global fit across both first-generation and non-first-generation alumni, with only a few local areas of misfit (see Table 2). Specifically, five areas of misfit were identified, four of which were associated with item 31. Additionally, for non-first-generation students, the model underestimated the relationship between items 21 and 22 (ICI subscale). Despite these minor local misfit issues, correlation residuals did not exceed 0.23, indicating an overall acceptable fit.

The bifactor model also fit the data well, and as expected, showed a statistically significant improvement over the five-factor model due to its additional parameters. However, statistical significance alone does not justify a more complex model. Parameter estimates for many items expected to represent the CMI were nonsignificant, suggesting that the CMI was not well-defined (see Table 3). Further, local areas of misfit from the five-factor model largely persisted in the bifactor model. Given the negligible substantive improvement and lack of theoretical support for the bifactor model, we selected the more interpretable five-factor model for further analysis.

Using this five-factor model, we proceeded to test measurement invariance across first-generation and non-first-generation alumni. The configural model fit the data well, $\chi^2(530) = 7{,}163.63$, $p < 0.01$, SRMR = 0.04, RMSEA = 0.05, CFI = 0.95 (see Table 2), supporting the consistency of the factor structure across both groups. Correlation residuals were identical to those from the sample-specific analyses. These findings provided a strong foundation for further invariance testing, ensuring that comparisons between first-generation and non-first-generation alumni would be meaningful and not driven by measurement inconsistencies.

---

[5] Skewness values greater than |2| and kurtosis values greater than |7| are considered indicative of univariate non-normality (Finney & DiStefano, 2013).

[6] The Satorra-Bentler adjustment corrects chi-square values, fit indices, and standard errors to account for multivariate non-normality, ensuring more robust and accurate model evaluation.

Table 1. *Inter-Item Correlations and Descriptive Statistics of the NACM Items for Samples 1 & 2*

Item	17	29	32	19	18	8	9	10	11	31	24	33	24	16	20	25	23	26	28	30	12	13	14	21	22
17	-	0.60	0.68	0.52	0.60	0.42	0.39	0.34	0.43	0.28	0.46	0.18	0.19	0.22	0.18	0.18	0.16	0.13	0.12	0.54	0.54	0.53	0.55	0.44	0.39
29	0.62	-	0.56	0.46	0.53	0.38	0.35	0.31	0.39	0.30	0.42	0.18	0.18	0.21	0.14	0.16	0.15	0.11	0.11	0.53	0.56	0.54	0.54	0.50	0.40
32	0.68	0.56	-	0.46	0.49	0.38	0.35	0.30	0.39	0.29	0.42	0.18	0.20	0.22	0.13	0.14	0.12	0.07	0.07	0.47	0.48	0.47	0.48	0.44	0.35
19	0.55	0.48	0.48	-	0.56	0.38	0.36	0.32	0.39	0.35	0.41	0.22	0.22	0.26	0.15	0.17	0.18	0.14	0.11	0.47	0.44	0.42	0.44	0.37	0.37
18	0.58	0.53	0.49	0.55	-	0.43	0.40	0.37	0.43	0.31	0.42	0.18	0.18	0.22	0.16	0.16	0.17	0.11	0.10	0.50	0.51	0.48	0.51	0.41	0.40
8	0.42	0.40	0.39	0.39	0.41	-	0.82	0.66	0.77	0.50	0.69	0.30	0.29	0.33	0.20	0.19	0.19	0.15	0.14	0.35	0.36	0.33	0.35	0.29	0.27
9	0.39	0.36	0.35	0.36	0.38	0.83	-	0.66	0.74	0.47	0.66	0.27	0.27	0.30	0.21	0.18	0.19	0.15	0.14	0.33	0.33	0.32	0.32	0.26	0.25
10	0.33	0.32	0.31	0.32	0.34	0.66	0.67	-	0.64	0.44	0.59	0.27	0.25	0.29	0.17	0.16	0.18	0.14	0.11	0.30	0.31	0.29	0.29	0.25	0.22
11	0.43	0.40	0.40	0.41	0.42	0.78	0.75	0.65	-	0.51	0.75	0.31	0.30	0.34	0.20	0.20	0.20	0.16	0.14	0.37	0.37	0.35	0.36	0.31	0.27
31	0.29	0.31	0.29	0.35	0.30	0.50	0.47	0.45	0.51	-	0.49	0.38	0.32	0.37	0.08	0.10	0.23	0.12	0.06	0.26	0.25	0.21	0.21	0.24	0.21
24	0.45	0.42	0.44	0.41	0.41	0.71	0.67	0.60	0.76	0.51	-	0.33	0.32	0.36	0.18	0.21	0.20	0.15	0.14	0.39	0.39	0.38	0.39	0.32	0.29
33	0.19	0.19	0.19	0.24	0.17	0.31	0.29	0.27	0.32	0.38	0.34	-	0.69	0.74	0.05	0.11	0.14	0.10	0.07	0.18	0.17	0.17	0.16	0.16	0.14
27	0.20	0.18	0.20	0.23	0.19	0.29	0.28	0.25	0.31	0.32	0.32	0.70	-	0.80	0.05	0.11	0.11	0.08	0.06	0.18	0.16	0.16	0.16	0.16	0.15
16	0.24	0.21	0.22	0.26	0.22	0.34	0.32	0.29	0.35	0.37	0.37	0.75	0.81	-	0.05	0.11	0.12	0.08	0.06	0.21	0.20	0.19	0.19	0.19	0.17
20	0.17	0.17	0.13	0.13	0.16	0.22	0.21	0.18	0.21	0.09	0.20	0.05	0.06	0.06	-	0.47	0.37	0.57	0.68	0.17	0.16	0.16	0.16	0.14	0.17
25	0.17	0.17	0.13	0.15	0.15	0.20	0.19	0.18	0.20	0.10	0.19	0.10	0.11	0.11	0.47	-	0.26	0.41	0.48	0.21	0.17	0.20	0.19	0.16	0.19
23	0.15	0.17	0.12	0.19	0.15	0.21	0.19	0.18	0.21	0.26	0.20	0.15	0.10	0.13	0.38	0.28	-	0.57	0.42	0.15	0.12	0.09	0.11	0.14	0.17
26	0.13	0.15	0.09	0.13	0.12	0.19	0.18	0.16	0.19	0.14	0.18	0.11	0.09	0.10	0.58	0.44	0.57	-	0.72	0.13	0.11	0.10	0.11	0.11	0.14
28	0.12	0.13	0.08	0.11	0.11	0.17	0.16	0.14	0.16	0.07	0.16	0.08	0.08	0.07	0.69	0.50	0.41	0.72	-	0.12	0.11	0.11	0.12	0.11	0.13
30	0.54	0.54	0.49	0.49	0.50	0.35	0.32	0.27	0.36	0.27	0.38	0.18	0.18	0.20	0.18	0.20	0.17	0.16	0.14	-	0.61	0.63	0.62	0.56	0.49
12	0.54	0.58	0.50	0.45	0.51	0.37	0.34	0.30	0.37	0.25	0.39	0.17	0.18	0.21	0.16	0.18	0.13	0.13	0.12	0.61	-	0.71	0.77	0.60	0.50
13	0.54	0.57	0.49	0.44	0.49	0.35	0.32	0.29	0.36	0.23	0.39	0.18	0.17	0.19	0.17	0.20	0.12	0.13	0.13	0.63	0.71	-	0.71	0.55	0.47
14	0.56	0.57	0.50	0.45	0.52	0.38	0.34	0.29	0.38	0.23	0.40	0.17	0.16	0.19	0.18	0.18	0.12	0.13	0.12	0.62	0.76	0.71	-	0.57	0.49
21	0.45	0.51	0.44	0.37	0.41	0.31	0.29	0.25	0.32	0.27	0.34	0.16	0.17	0.20	0.14	0.17	0.16	0.14	0.12	0.57	0.61	0.58	0.57	-	0.61
22	0.39	0.41	0.34	0.34	0.38	0.26	0.24	0.23	0.28	0.23	0.29	0.14	0.15	0.17	0.18	0.19	0.18	0.18	0.15	0.50	0.51	0.49	0.48	0.60	-

Item	17	29	32	19	18	8	9	10	11	31	24	33	24	16	20	25	23	26	28	30	12	13	14	21	22
Phase 1 Sample (*N* = 8,241)																									
M	3.75	3.42	3.83	3.44	4.02	3.99	4.00	4.01	3.90	4.07	3.76	3.15	3.80	3.64	3.49	2.77	3.20	3.02	2.99	3.12	3.28	2.97	3.26	3.03	2.90
SD	1.06	1.19	1.13	1.19	0.93	0.94	0.97	0.90	0.98	1.05	1.08	1.40	1.17	1.25	1.02	1.04	1.17	1.13	1.10	1.13	1.15	1.20	1.16	1.20	1.12
Skewness	-0.98	-0.53	-1.04	-0.53	-1.24	-1.25	-1.22	-1.19	-1.07	-1.28	-0.87	-0.01	-0.71	-0.56	-0.53	0.25	-0.30	0.00	0.08	-0.21	-0.42	-0.05	-0.40	-0.10	0.02
Kurtosis	0.43	-0.69	0.34	-0.64	1.67	1.69	1.45	1.67	0.94	1.14	0.16	-1.37	-0.49	-0.80	-0.34	-0.59	-0.94	-0.95	-0.93	-0.77	-0.79	-1.02	-0.81	-1.00	-0.82
Phase 2 Sample (*N* = 8,240)																									
M	3.75	3.43	3.84	3.43	4.02	4.00	4.00	4.00	3.90	4.06	3.77	3.13	3.81	3.64	3.50	2.77	3.21	3.04	3.00	3.12	3.28	2.98	3.26	3.03	2.89
SD	1.07	1.16	1.13	1.19	0.93	0.94	0.97	0.91	0.98	1.04	1.08	1.38	1.17	1.24	1.01	1.05	1.17	1.11	1.10	1.11	1.15	1.19	1.16	1.20	1.11
Skewness	-0.98	-0.55	-1.02	-0.50	-1.25	-1.26	-1.20	-1.18	-1.09	-1.27	-0.86	0.01	-0.76	-0.56	-0.50	0.24	-0.25	-0.01	0.08	-0.24	-0.45	-0.05	-0.42	-0.09	0.03
Kurtosis	0.53	-0.59	0.31	-0.67	1.74	1.72	1.35	1.54	0.96	1.14	0.14	-1.36	-0.39	-0.80	-0.37	-0.61	-0.97	-0.93	-0.91	-0.73	-0.74	-1.03	-0.78	-1.02	-0.81

Note. Sample 1 correlations are below the diagonal while Sample 2 correlations are above the diagonal.

Table 2. *Fit Indices and Difference Tests of Hypothesized Models using Sample 1 (N = 8,241)*

Model	χ^2	df	SRMR	RMSEA	CFI	Local misfit	$\Delta\chi^2$
First-Gen (FG; *N* = 2,537)							
Model 1: Five-Factor	2,284.73*	265	0.04	0.06	0.95	Q31 □ Q33, Q27, Q16, Q23	
Model 2: Bifactor	1,807.20*	251	0.04	0.05	0.96	Q31 □ Q23	477.53*
Non-First-Gen (NFG; *N* = 5,704)							
Model 1: Five-Factor	4,878.90*	265	0.04	0.06	0.95	Q31 □ Q33, Q16, Q23; Q21 □ Q22	
Model 2: Bifactor	4,026.02*	251	0.04	0.05	0.96	Q31 □ Q33, Q16, Q23; Q21 □ Q22	852.88*
Measurement Invariance							
Five-Factor: Configural	7,163.63*	530	0.04	0.05	0.95	FG: Q31 □ Q33, Q27, Q16, Q23 NFG: Q31 □ Q33, Q16, Q23; Q21 □ Q22	
Five-Factor: Metric	7,303.30*	550	0.04	0.05	0.95	FG: Q31 □ Q33, Q27, Q16, Q23 NFG: Q31 □ Q33, Q16, Q23; Q21 □ Q22	139.67*
Five-Factor: Scalar	7,518.17*	570	0.04	0.05	0.95	FG: Q31 □ Q33, Q27, Q16, Q23 NFG: Q31 □ Q33, Q16, Q23; Q21 □ Q22	214.88*

Note. $^*p < .01$. □ = the item has an underrepresented relationship with the following items. For example, for the first-gen five-factor model, the relationship between item 31 and items 33, 27, and 16 is underrepresented

Table 3. *Unstandardized (Standardized) Path Coefficients and Unstandardized Error Variances for Five-Factor and Bifactor Models using Sample 1*

Item	First-Generation							Non-First-Generation						
	CPP	CS	EM	CE	ICI	CMI	Error Variance	CPP	CS	EM	CE	ICI	CMI	Error Variance
							Five-Factor Model							
17	0.95 (0.84)						0.37	0.83 (0.81)						0.36
29	0.94 (0.76)						0.64	0.88 (0.76)						0.57
32	0.90 (0.75)						0.64	0.82 (0.75)						0.53
19	0.86 (0.70)						0.79	0.78 (0.66)						0.77
18	0.73 (0.73)						0.47	0.64 (0.71)						0.40
8		0.93 (0.91)					0.18		0.80 (0.90)					0.16
9		0.92 (0.87)					0.26		0.81 (0.87)					0.21
10		0.74 (0.76)					0.39		0.63 (0.73)					0.35
11		0.94 (0.88)					0.26		0.83 (0.88)					0.20
31		0.69 (0.61)					0.80		0.56 (0.56)					0.69
24		0.96 (0.83)					0.43		0.84 (0.81)					0.37
33			1.08 (0.82)				0.58			1.05 (0.78)				0.70
27			0.82 (0.81)				0.34			1.04 (0.87)				0.34
16			1.03 (0.93)				0.17			1.15 (0.93)				0.22
20				0.82 (0.77)			0.45				0.75 (0.75)			0.43
25				0.64 (0.59)			0.74				0.58 (0.56)			0.73
23				0.70 (0.60)			0.88				0.62 (0.53)			0.99
26				0.95 (0.83)			0.41				0.91 (0.82)			0.41
28				0.97 (0.86)			0.32				0.95 (0.87)			0.30
30					0.88 (0.77)		0.54					0.84 (0.75)		0.54
12					1.02 (0.86)		0.36					0.97 (0.86)		0.36
13					1.03 (0.84)		0.46					0.97 (0.82)		0.45
14					1.02 (0.85)		0.39					0.96 (0.85)		0.38
21					0.89 (0.74)		0.65					0.85 (0.71)		0.71
22					0.71 (0.64)		0.72					0.69 (0.62)		0.78

Item	First-Generation							Non-First-Generation						
	CPP	CS	EM	CE	ICI	CMI	Error Variance	CPP	CS	EM	CE	ICI	CMI	Error Variance
Bifactor Model														
17	0.95 (0.84)					0.01 (0.01)	0.38	0.84 (0.81)					0.01 (0.01)	0.36
29	0.94 (0.76)					0.04 (0.03)	0.64	0.88 (0.76)					0.01 (0.01)	0.57
32	0.90 (0.75)					0.04 (0.03)	0.64	0.82 (0.75)					0.05 (0.05)	0.53
19	0.85 (0.69)					0.22 (0.18)	0.75	0.78 (0.66)					-0.04(-0.03)	0.77
18	0.72 (0.73)					0.03 (0.03)	0.47	0.64 (0.71)					-0.06(-0.06)	0.40
8		0.93 (0.91)				-0.17 (-0.16)	0.16		0.72 (0.80)				-0.36(-0.41)	0.15
9		0.92 (0.87)				-0.23 (-0.22)	0.22		0.70 (0.75)				-0.48(-0.52)	0.14
10		0.74 (0.76)				-0.06 (-0.06)	0.40		0.58 (0.67)				-0.25(-0.28)	0.35
11		0.93 (0.87)				-0.02 (-0.02)	0.27		0.81 (0.85)				-0.19(-0.20)	0.20
31		0.72 (0.64)				0.44 (0.39)	0.55		0.57 (0.57)				-0.05(-0.05)	0.67
24		0.96 (0.83)				0.08 (0.07)	0.42		0.90 (0.87)				0.00 (0.00)	0.26
33			0.97 (0.74)			0.50 (0.38)	0.54			1.05 (0.78)			0.01 (0.01)	0.69
27			0.77 (0.77)			0.26 (0.26)	0.34			1.04 (0.87)			0.05 (0.04)	0.34
16			0.97 (0.87)			0.33 (0.30)	0.19			1.15 (0.93)			0.03 (0.02)	0.22
20				0.82 (0.77)			0.45				0.75 (0.75)			0.43
25				0.64 (0.59)			0.74				0.58 (0.56)			0.73
23				0.70 (0.60)			0.88				0.62 (0.53)			0.99
26				0.95 (0.83)			0.41				0.92 (0.82)			0.41
28				0.97 (0.86)			0.33				0.95 (0.87)			0.30
30					0.88 (0.77)		0.54					0.84 (0.75)		0.54
12					1.02 (0.86)		0.36					0.97 (0.85)		0.36
13					1.03 (0.84)		0.46					0.97 (0.82)		0.45
14					1.02 (0.85)		0.39					0.96 (0.84)		0.38
21					0.89 (0.74)		0.65					0.85 (0.71)		0.71
22					0.71 (0.64)		0.72					0.69 (0.62)		0.78

Note. R^2 can be calculated by squaring the standardized path coefficients. R^2 can be interpreted as the percentage of the item's variance that can be explained by its corresponding factor. For example, for item 31 for first-generation students, 37% of the item's variance can be explained by the CS factor.

Metric Invariance for the Five-Factor Model

The metric model, which constrained factor loadings to be equal across groups, fit the data well, $\chi^2(550) = 7{,}303.30$, SRMR = 0.04, RMSEA = 0.05, CFI = 0.95, with minimal local misfit (see Table 2). The identified areas of misfit mirrored those in the configural model, and correlation residuals did not exceed 0.25. While the $\Delta\chi^2$ test indicated a statistically significant difference between the configural and metric models, the difference in global and local fit was negligible. Item 27 had a larger factor loading difference between first-generation (0.82) and non-first-generation (1.04) alumni than other items, prompting further investigation. Scalar invariance testing showed that latent mean differences remained nearly identical whether item 27's parameters were constrained or freely estimated, differing only at the second decimal place (Kopp & Finney, 2013). Thus, we concluded that metric invariance was supported.

Scalar Invariance for the Five-Factor Model

The scalar model, which constrained item intercepts to be equal across groups, fit the data well, $\chi^2(570) = 7{,}518.17$, SRMR = 0.04, RMSEA = 0.05, CFI = 0.95. Local misfit areas mirrored those found in the metric model, with correlation residuals not exceeding 0.25. To assess misfit in reproducing item-level means, we examined mean residuals and found minimal issues. The largest mean residual (0.11 for item 33 in the first-generation group) was negligible, given the 1 to 5 Likert scale. The observed mean for item 33 was 3.76, whereas the model-implied mean was 3.65 under equality constraints, a negligible difference. Although the $\Delta\chi^2$ test showed a statistically significant difference between the metric and scalar models, this result was likely influenced by sample size. Given the negligible difference in global and local fit (see Table 2), scalar invariance was supported.

Latent mean differences, representing differences between first-generation and non-first-generation alumni on each subscale, were examined to determine substantive differences (see Table 6). The largest difference (0.66) was found on the EM subscale, with first-generation alumni scoring higher.

When examining the results from this championed five-factor invariant model, it appears that the NACM may function better without item 31. The item better represented EM than CS in Sample 1. Additionally, item 31 did not have a large R^2 value, indicating that it contributed little to the model's overall explanatory power (i.e., it is not "pulling its weight" in the model; see Table 3). Given these findings, we recommend removing item 31 from the scale to improve model clarity and interpretability.

In sum, the five-factor model was supported over the bifactor model, which failed due to an ill-defined CMI factor. However, item 31 continued to show notable misfit, with large correlation residuals involving items 33, 27, 16 (EM), and 23 (CE). Thus, we examined whether the misfit associated with item 31 was replicated using sample 2. If so, we could estimate measurement invariance after removing item 31 and evaluate if students of different first-generation statuses differed on NACM factors.

Sample 2

Replication

The bifactor and five-factor models fit the data well for first-generation and non-first-generation alumni using sample 2 data. Substantive results and decisions were identical to those when modeling sample 1 data. The misfit found using sample 1 was replicated using sample 2: item 31 continued to cause issues for both models. Thus, the five-factor model was championed for the above-mentioned reasons, and fit was assessed when removing item 31.

Configural Invariance of Five-Factor Model and Bifactor Model without Item 31

As expected, both models fit well globally when item 31 was removed (see Table 4), with fewer local misfit issues compared to models that included it. The five-factor model showed three areas of local misfit, with two correlation residuals involving item 25 (the relations with items 30 and 22 were underestimated by .16 and .15, respectively, for first-generation students). The bifactor model had two areas of local misfit, including the underestimated relationship between item 25 and item 30 for first-generation students (.16). Additionally, for both first-generation and non-first-generation students, the relationship between items 21 and 22 remained underestimated (.15 and .17, respectively). Removing item 31 resolved many local fit issues observed in Sample 1, though two new correlation residuals associated with item 25 emerged in Sample 2. Overall, the adjusted five-factor model in Sample 2 had fewer correlation residuals (three) than the five-factor model in Sample 1 with item 31 included (five), indicating an improvement in model fit.

A statistically significant $\Delta\chi^2$ test indicated that the bifactor model fit better than the five-factor model; however, this result was expected given the large sample size. Despite the statistical significance, the CMI factor remained poorly defined, with most items failing to demonstrate significant relationships with the CMI factor. Consequently, the five-factor model was reaffirmed as the preferred structure and used for subsequent invariance testing.

Parameter estimates for the five-factor configural model (see Table 5) showed that most NACM items had at least 50% of their variance explained by their respective factors. Reliability estimates (omega) were high for the subscales for both groups: for first-generation students, CPP ω = .85, CS ω = .93, EM ω = .88, CE ω = .85, and ICI ω = .90; for non-first-generation students, CPP ω = .87, EM ω = .89, CE ω = .86, ICI ω = .90, and CS ω = .94. These values suggest minimal measurement error in the observed scores, further supporting the robustness of the five-factor model without item 31.

Metric Invariance for the Five-Factor Model without Item 31

The metric model fit the data well globally, $\chi^2(503)$ = 6,426.30, SRMR = 0.04, RMSEA = 0.05, CFI = 0.95, with three areas of local misfit (see Table 4). The $\Delta\chi^2$ test indicated a statistically significant difference between the configural and metric models, but the difference in global and local fit was negligible. Items 27 and 16 continued to show larger factor loading differences between first-generation (0.80 and 0.96) and non-first-generation (1.01 and 1.14) alumni, as observed in Sample 1. However, latent means were minimally different when tested under constrained and freely estimated conditions, supporting metric invariance across groups.

Scalar Invariance

The scalar model fit the data well globally and had few local issues, $\chi^2(522) = 6{,}637.88$, SRMR = 0.04, RMSEA = 0.05, CFI = 0.95. As found using sample 1, the difference between the metric and scalar models was negligible when examining the global and local fit indices.

Scalar Invariance & Latent Means Modeling. Given the equivalence of factor structures, unstandardized factor coefficients, and item intercepts across first-generation and non-first-generation students, latent mean differences on the five constructs were computed. The Cohen's *d* effect sizes for four out of the five subscales were small, indicating small differences across first-generation status (< .15; Table 6). However, the latent effect size for EM was large (.72). The first-generation sample was 0.72 standard deviations higher in EM than the non-first-generation sample. When examining the observed means and effect sizes (Table 6), they tell a similar story: first-generation alumni are scoring higher on the EM subscale compared to their peers (.72). The observed and latent effect sizes are similar because not only was measurement invariance established, but reliability (e.g., omega) was high.

Prior to this study, it was unclear if the observed differences across the two groups were trustworthy. Now, with evidence of measurement invariance, these observed subscale differences can be trusted. Given the wording of the items, it makes sense that first-generation students would report higher on EM. Since measurement invariance was established, the differences in EM do not reflect bias. It is therefore appropriate to compare these two groups of alumni on the five NACM subscales at the observed or latent level.

Discussion

Students expect college to help them find and sustain a productive career (Blaich & Wise, 2021). One way to collect evidence associated with this expectation is to survey alumni about their career outcomes. The NACM survey was created for this purpose and is widely used, yet prior to this study, it was unclear how to score the survey and whether comparisons across groups were justified.

Our study explored the scoring of the NACM, yielding surprisingly favorable results. This was unexpected given the NACM's limited theoretical foundation; specifically, the rationale for how items were selected and aligned with each construct had not been clearly documented. We found evidence supporting the five constructs purportedly measured by the NACM, providing empirical justification for using five subscale scores. We also found scalar invariance—the most stringent type of invariance—upheld between first-generation and non-first-generation alumni, meaning that NACM scores can be meaningfully compared across these groups. Recall we predicted that we might observe non-invariance for the CPP, ICI, and EM subscales. This was not the case – measurement invariance was upheld (a positive outcome) for all subscales!

Table 4. *Fit Indices and Difference Tests of Hypothesized Models without Item 31 using Sample 2 (N = 8,240)*

Model	χ^2	df	SRMR	RMSEA	CFI	Local misfit	$\Delta\chi^2$
First-Gen (FG; *N* = 2,537)							
Model 1: Five-Factor	2,052.75*	242	0.04	0.05	0.95	Q25 □ Q30, Q22; Q21 □ Q22	389.17*
Model 2: Bifactor	1,663.58*	229	0.04	0.05	0.96	Q25 □ Q30; Q21 □ Q22	
Non-First-Gen (NFG; *N* = 5,703)							
Model 1: Five-Factor	4,271.39*	242	0.04	0.05	0.95	Q21 □ Q22	885.55*
Model 2: Bifactor	3,385.84*	229	0.04	0.05	0.96	Q21 □ Q22	
Measurement Invariance							
Five-Factor: Configural	6,324.14*	484	0.04	0.05	0.95	FG: Q25 □ Q30, Q22; Q21 □ Q22 NFG: Q21 □ Q22	
Five-Factor: Metric	3,426.30*	503	0.04	0.05	0.95	FG: Q25 □ Q30, Q22; Q21 □ Q22 NFG: Q21 □ Q22	102.16*
Five-Factor: Scalar	6,637.88*	522	0.04	0.05	0.95	FG: Q25 □ Q30, Q22; Q21 □ Q22 NFG: Q21 □ Q22	211.58*

Note. **p* < .01. □ = the item has an underrepresented relationship with the following items. For example, for the first-gen five-factor model, the relationship between item 31 and items 33, 27, and 16 is underrepresented.

Table 5. *Unstandardized (Standardized) Path Coefficients and Unstandardized Error Variances for Five-Factor Model Using Sample 2 Without Item 31*

Item	Factor Pattern Coefficients											
	First-Generation						Non-First-Generation					
	CPP	CS	EM	CE	ICI	Error Variance	CPP	CS	EM	CE	ICI	Error Variance
17	0.92 (0.83)					0.40	0.92 (0.81)					0.37
29	0.88 (.075)					0.61	0.92 (0.74)					0.60
32	0.87 (0.73)					0.66	0.88 (0.73)					0.56
19	0.83 (0.67)					0.84	0.83 (0.66)					0.78
18	0.72 (0.74)					0.43	0.72 (0.74)					0.38
8		0.92 (0.91)				0.19		0.92 (0.89)				0.17
9		0.90 (0.87)				0.25		0.93 (0.87)				0.21
10		0.74 (0.76)				0.39		0.73 (0.73)				0.36
11		0.93 (0.87)				0.26		0.94 (0.87)				0.22
24		0.93 (0.81)				0.45		0.95 (0.80)				0.40
33			1.02 (0.80)			0.60			1.02 (0.78)			0.68
27			0.80 (0.83)			0.30			1.01 (0.86)			0.38
16			0.96 (0.92)			0.18			1.14 (0.92)			0.24
20				0.81 (0.79)		0.41				0.81 (0.74)		0.44
25				0.59 (0.56)		0.79				0.63 (0.55)		0.76
23				0.68 (0.58)		0.91				0.69 (0.54)		0.96
26				0.91 (0.81)		0.44				0.99 (0.82)		0.40
28				0.97 (0.87)		0.31				1.05 (0.87)		0.29
30					0.84 (0.76)	0.53					0.84 (0.75)	0.53
12					1.00 (0.86)	0.35					0.99 (0.86)	0.35
13					1.00 (0.83)	0.45					0.97 (0.81)	0.47
14					1.01 (0.85)	0.38					0.99 (0.85)	0.36
21					0.88 (0.73)	0.69					0.85 (0.70)	0.72
22					0.69 (0.63)	0.74					0.69 (0.62)	0.76

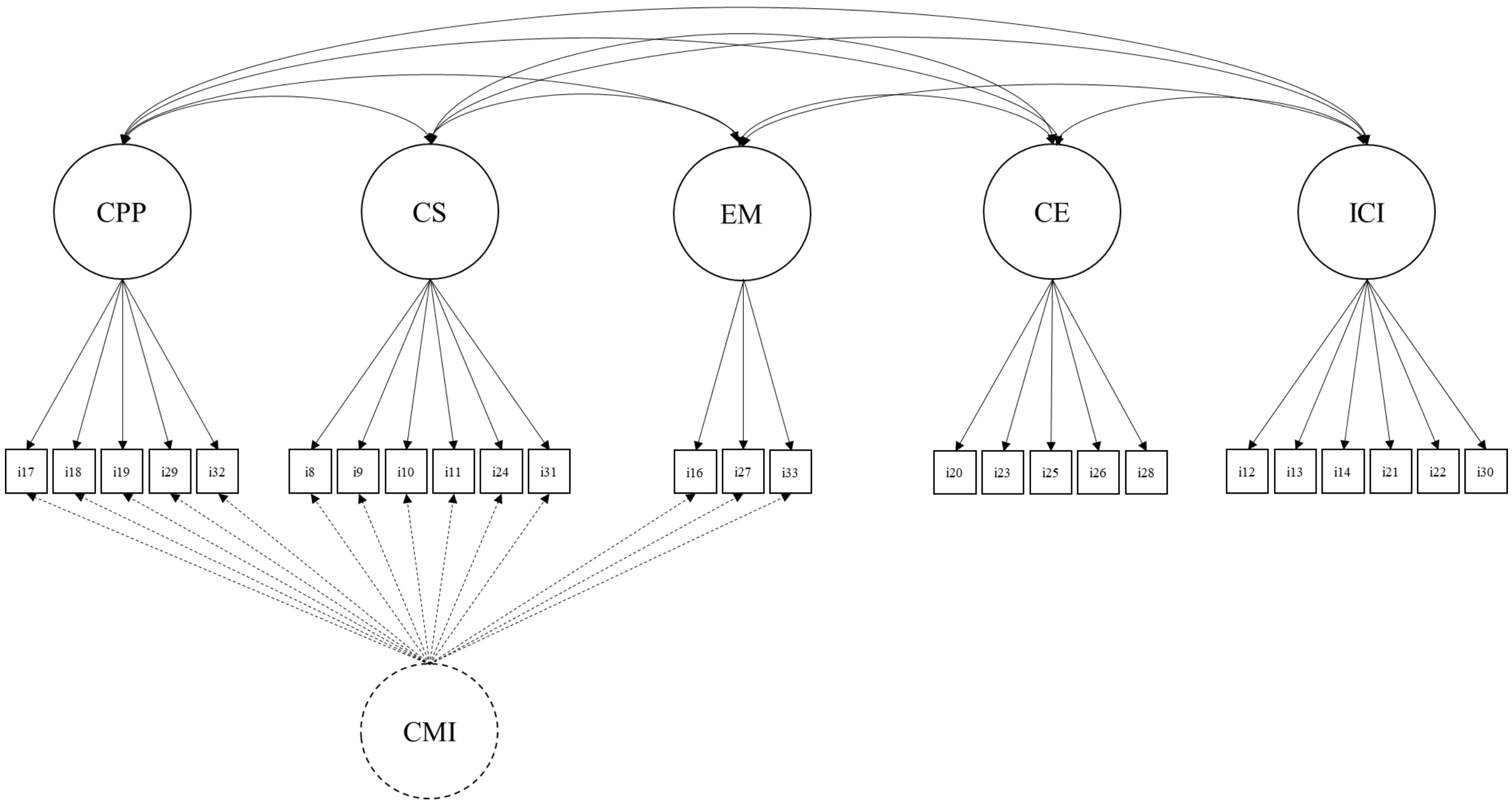
Note. All values are significant at $p < .01$.

Table 6. *Observed and Latent NACM Factor Means (Standard Deviations) and Effect Sizes*

Sample	CPP	CS	EM	CE	ICI
Observed Full Sample with Item 31					
First-Gen	3.61 (.93)	3.87 (.89)	4.05 (1.01)	3.06 (.87)	3.04 (.96)
Non-First-Gen	3.73 (.85)	4.00 (.78)	3.30 (1.14)	3.11 (.83)	3.12 (.93)
Observed Cohen's *d*	.14	.16	.68	.06	.09
Sample 1 with Item 31					
First-Gen	3.61 (.84)	3.84 (.87)	4.02 (.92)	3.08 (.76)	3.04 (.89)
Non-First-Gen	3.73 (.75)	3.97 (.76)	3.32 (1.04)	3.10 (.72)	3.12 (.84)
Latent Cohen's *d*	.14	.16	.66	.03	.08
Sample 2 with Item 31					
First-Gen	3.61 (.80)	3.85 (.86)	4.04 (.87)	3.07 (.74)	3.05 (.87)
Non-First-Gen	3.73 (.74)	3.97 (.76)	3.30 (1.04)	3.12 (.71)	3.11 (.84)
Latent Cohen's *d*	.14	.15	.72	.07	.07
Sample 2 without Item 31					
First-Gen	3.61 (.80)	3.85 (.87)	4.04 (.87)	3.07 (.74)	3.05 (.87)
Non-First-Gen	3.73 (.74)	3.97 (.76)	3.30 (1.04)	3.12 (.71)	3.11 (.84)
Latent Cohen's *d*	.14	.15	.72	.07	.07

Note. Bolded values indicate significant and non-ignorable differences across the groups. Cohen's *d* was computed using the pooled latent standard deviation across groups since homogeneity of variance (HOV) was established (Kline, 2023). HOV can be established by simply examining the variance estimates for the latent factors between groups. For example, an effect size of 0.72 indicates first-generation alumni, on average, scored 0.72 standard deviations higher on EM compared to non-first-generation alumni.

Figure 1. Five-Factor and Bifactor NACM Structures



Note. The five-factor model does not include the dashed lines. CPP = career pathway preparation, CS = career satisfaction, EM = economic mobility, CE = community engagement, ICI = institutional career investment, CMI = Career Mobility Index. In the bifactor model, the CPP, CS, and EM subscales would change to be specific factors since they cannot represent the same variance.

Indeed, given the strong evidence of measurement invariance, we could explore mean differences across college-generation status. Notably, we found a large difference between first-generation and non-first-generation alumni on the EM subscale (.72), with first-generation alumni reporting higher than non-first-generation alumni. This suggests that first-generation alumni may experience greater upward economic mobility than their non-first-generation peers.

Additional Psychometric Work

Based on our findings, we advocate for reporting five subscale scores (CPP, CS, EM, CE, and ICI) and recommend discontinuing the use of the CMI subscale score in future reporting. We did not find support for reporting a CMI subscale score. Additionally, our analysis revealed that one item on the NACM (31 – "My salary is enough to pay my bills every month") does not function well. Therefore, we recommend its removal from future versions of the CS subscale. Item 31 was associated with many areas of local misfit. Moreover, the item had a low R^2 (0.37 for first-generation and 0.31 for non-first-generation; see Table 3), accounting for less than half of the variance, across first-generation status. If item 31 is removed, additional psychometric studies are needed, as it changes the variance that is modeled. Specifically, the CS subscale needs to be reevaluated to ensure that it is clearly and theoretically defined without item 31.

If item 31 is retained, then scores should be interpreted with caution. Specifically, item 31 shows consistent misfit and explains little variance in CS. Thus, retaining item 31 could lead to inconsistencies within the CS subscale. Retaining item 31 can also complicate scoring and interpretation. The item's misalignment with the CS subscale can potentially skew results and make it difficult to draw meaningful conclusions. This finding should be communicated to universities that receive score reports to ensure they interpret and use scores appropriately, particularly in understanding how first-generation and non-first-generation alumni experience economic mobility differently. Institutions should consider these differences when making data-informed decisions about career services, financial support programs, and alumni engagement strategies to promote equitable opportunities for all graduates.

Implications

Reliable data on alumni career outcomes, beyond traditional measures such as employment status and salary, provides institutions with valuable insights. These broader outcomes can help institutions demonstrate the value of their educational experience. First, data on alumni career outcomes can offer a compelling narrative about how higher education prepares graduates for long-term success. By focusing on outcomes beyond salary, such as career preparation, satisfaction, and engagement, institutions gain a more comprehensive understanding of alumni success. Second, disaggregating the data (e.g., by first-generation status) allows institutions to identify and address equity gaps in career readiness and support. Finally, linking career outcomes to specific educational experiences creates a feedback loop that career preparation programs and career readiness education can use to better address student needs.

Our study supports score interpretations for the NACM, a tool that institutions can use to enhance career preparation programs and career readiness education (although it lacks a strong theoretical framework). For example, institutions with low CPP scores can use these scores to identify gaps in their career preparation programs, internship opportunities, or employer partnerships. Similarly, low ICI scores indicate areas for improvement in how institutions invest in career services or provide alumni networking opportunities. Importantly, these findings can help campuses frame the narrative around their return on investment more effectively. Using CPP and ICI data, institutions can demonstrate the tangible ways they prepare students for successful careers, which may improve perceptions of the value of higher education among stakeholders.

Subscales like EM, CS, and CE also provide a broader view of alumni success that aligns with evolving definitions of career outcomes. While the NACM addresses a complex construct in career mobility, our study provides psychometric evidence for scoring the survey. However, future research should address the NACM's theoretical limitations to further refine its value for institutions and career professionals.

Future Research

Future research on the NACM should begin by building a theoretical case for the subscales. A robust literature review supporting the five subscales is recommended to better link the items' empirical domain to their theoretical underpinnings. Given that reporting a profile of subscale scores is now possible, the NACM could provide valuable insights when combined with other datasets to answer a broader range of research questions. For example, researchers may conceptualize the five NACM factors as independent or dependent variables. In this context, one could study which college experiences predict these factors (i.e., subscales as dependent variables). Alternatively, the NACM subscales could be paired with other datasets to explore relationships with variables of interest, such as well-being or life satisfaction (e.g., subscales as predictors).

Limitations

It is important to acknowledge our study's limitations. A glaring limitation is the lack of theoretical evidence for the NACM constructs. Although the items were written by content experts, there is no underlying theory to support the subscales. Further, removing item 31 would change the CS scale. Thus, theoretical consideration should be given to the CS scale. Overall, all scores should be interpreted with caution.

Collecting data from alumni poses considerable challenges. A primary limitation is sampling bias. Alumni who respond to surveys may differ systematically from non-respondents, resulting in findings that may not accurately reflect the broader alumni population. Consequently, the results of this study may have limited applicability to all college alumni.

Additionally, our sample included a substantial amount of missing data. After removing records with missing data on the 25 NACM items and the first-generation status variable, the sample size decreased from 19,819 to 16,481. Missing data was deleted, which presents challenges (Enders, 2022). Future studies could employ better missing data techniques.

Furthermore, the sample used was relatively homogeneous, with most participants identifying as female and White. This homogeneity limits the generalizability of findings to other populations. Future research with a more diverse respondent pool could explore demographic differences beyond first-generation status, providing additional insights into how various groups experience and interpret career outcomes.

Conclusion

The NACM functions well as a tool for assessing alumni career outcomes, allowing for meaningful score comparisons between first-generation and non-first-generation alumni. However, we recommend discontinuing the scoring and reporting of the CMI, as our findings did not support its use. As demand for evidence of career success continues to grow, the NACM remains a valuable tool for measuring career outcomes. Ongoing psychometric work is essential to ensure that scores are interpreted equitably and with informed understanding. Higher education institutions need reliable post-graduation outcome data to demonstrate their value to students and stakeholders. Students seek assurance that their education will have a lasting impact on their career (Blaich & Wise, 2021). Therefore, continued research on the value of a college degree and refining measures like the NACM is critical.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bandalos, D. L., & Finney, S. J. (2019). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 98-122). Routledge. https://doi.org/10.4324/9781315755649

Blaich, C., & Wise, K. (2021). It's time to bring students into the conversation about student success. *Change: The Magazine of Higher Learning*, *53*(6), 4–11. https://doi.org/10.1080/00091383.2021.1987786

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, *17*(1), 10–17. https://doi.org/10.1111/j.1745-3992.1998.tb00616.x

Büchel, F., & Mertens, A. (2004). Overeducation, undereducation, and the theory of career mobility. *Applied Economics*, *36*(8), 803–816. https://doi.org/10.1080/0003684042000229532

The Career Leadership Collective. (2021a). Career mobility index factor analysis [Unpublished internal report].

The Career Leadership Collective. (2021b). NACM factor analysis [Unpublished internal report].

The Career Leadership Collective. (2022). *National alumni career mobility annual report: Transforming career practices to increase ROI and equitable career success*. https://www.datocms-assets.com/62658/1695854082-nacm_annual_report_sept2022.pdf

Chien, C.-L., Montjouridès, P., & van der Pol, H. (2016). Global trends of access to and equity in postsecondary education. In A. Mountford-Zimdars & N. Harrison (Eds.), *Access to Higher Education*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315684574-7/global-trends-access-equity-postsecondary-education-chiao-ling-chien-patrick-montjourid%C3%A8s-hendrik

Choy, S. P. (2001). Students whose parents did not go to college: Postsecondary access, persistence, and attainment. In J. Wirt, S. Choy, D. Gerald, S. Provasnik, P. Rooney, S. Watanabe, R. Tobin, & M. Glander (Eds.), *The condition of education 2001* (NCES 2001–072, pp. xviii–xliii). U.S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/pubs2001/2001072.pdf

Dumford, A. D., & Miller, A. L. (2017). Assessing alumni success: Income is NOT the only outcome! *Assessment & Evaluation in Higher Education*, *42*(2), 195–207. https://doi.org/10.1080/02602938.2015.1098587

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). The Guilford Press.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *A second course in structural equation modeling* (2nd ed., pp. 439–492). Information Age.

Finney, S. J., DiStefano, C., & Kopp, J. (2016). Overview of estimation methods and preconditions for their application with structural equation modeling. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 135–165). Hogrefe.

Gable, R. (2021). *The hidden curriculum: First generation students at legacy universities*. Princeton University Press.

Hamilton, I. (2023, June 13). *56% of all undergraduates are first-generation college students*. Forbes Advisor. https://www.forbes.com/advisor/education/online-colleges/first-generation-college-students-by-state/

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th edition). Guilford Press.

Kopp, J. P., & Finney, S. J. (2013). Linking academic entitlement and student incivility using latent means modeling. *The Journal of Experimental Education*, *81*(3), 322–336. https://doi.org/10.1080/00220973.2012.727887

Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature*. Commissioned report for the National Symposium on Postsecondary Student Success. https://nces.ed.gov/npec/pdf/Kuh_Team_Report.pdf

Manzoni, A., & Streib, J. (2019). The equalizing power of a college degree for first-generation college students: Disparities across institutions, majors, and achievement levels. *Research in Higher Education*, *60*(5), 577–605. https://doi.org/10.1007/s11162-018-9523-1

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, *92*(6), 1086–1120. https://doi.org/10.1086/261276

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *The Journal of Higher Education*, *75*(3), 249–284. https://www.jstor.org/stable/3838816

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc. https://psycnet.apa.org/record/1996-97111-016

Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy*, *98*(1), 169–192. https://www.jstor.org/stable/2937647

Widaman, K. F., & Olivera-Aguilar, M. (2023). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (2nd ed., pp. 367–384). Guilford Publications.

How to cite this article: Jones, M. R., Good, M. R., & Finney, S. J. (2025). Measuring alumni career outcomes: A validity study. *Journal of Student Affairs Inquiry, Improvement, and Impact*, *8*(1), 48–72. https://doi.org/10.18060/29219

Editorial Acknowledgement. The *Journal of Inquiry, Improvement, and Impact* (*JSAIII*) uses an open and transparent peer-review process designed to foster professional, rigorous dialogue that strengthens each article. This approach also enables us to publicly recognize the essential contributions of our editorial team—work that is often invisible and uncredited. By acknowledging these efforts, we aim to bring greater equity to scholarly publishing and highlight the collaborative labor behind every published piece. The following individuals contributed to the editorial development of this article.

Kimberly Yousey-Elsener, Ph.D.
Guest Editor
Binghamton University
Visiting Research Associate Professor, Higher Education and Student Affairs
LinkedIn | ORCiD

Anna Mroch, M.S.E.
Assistant Editor
William & Mary
Assistant Vice President for Planning, Strategy & Integrative Practices and Director of Assessment
LinkedIn

Production editing for this article was completed by *JSAIII* Editor, Heather J. Strine-Patterson, as part of the journal's commitment to publishing quality scholarship with care and consistency.

Appendix
National Alumni Center Mobility Survey

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1	2	3	4	5

Career Pathway Preparation

17. My bachelor's degree helped prepare me for my career.
29. I was well prepared to begin my career when I graduated.
32. My bachelor's degree helped me get started in my career.
18. I am satisfied with my educational experience for my bachelor's degree.
19. My bachelor's degree is worth the tuition I paid.

Career Satisfaction

8. I am satisfied with my career so far.
9. I am satisfied with my current career.
10. I am satisfied with the level of responsibility of my current job.
11. I am content with the progression of my career.
24. I have advanced in my career as I had hoped.
31. My salary is enough to pay my bills every month.

Economic Mobility

33. I currently earn a higher income than the households in which I grew up.
27. I expect to earn more in the future than the household in which I grew up.
16. My earning potential is higher than the household in which I grew up.

Community Engagement

20. I am involved in my community.
25. I have received recognition for my community involvement.
23. I regularly donate money to charitable causes.
26. I regularly donate time to charitable causes.
28. I volunteer in my community regularly.

Institutional Career Investment

30. My institution invested in my career.
12. My institution helped me to understand career opportunities.
13. My institution helped me create a plan for my career.
14. My institution helped me envision my career options.
21. My institution helped me to network with employers.
22. My institution helped me to network with alumni.

Note. Items were taken from The Career Leadership Collective (CLC) branded, publicly available factor sheets and annual report cited in the References section (CLC, 2021a, 2021b, 2022).