# Understanding Student Interest in Substance Use Disorder Counseling: A Mixed-Methods Approach Using Machine Learning and Qualitative Interviews

Erika Hernandez Acton, Linda Reynolds, Bryan Acton, Rory Eckhardt, Giovanni Scaringi, and Nadine Mastroleo

Binghamton University

# Author Note

Correspondence concerning this article should be addressed to Erika Hernandez Acton, Binghamton University, Email: [erika@binghamton.edu](mailto:erika@binghamton.edu)

# Abstract

*Keywords*: substance use disorder, counseling, workforce development, student interest, career choice, machine learning, random forest, qualitative interviews, mixed-methods

# Understanding Student Interest in Substance Use Disorder Counseling: A Mixed-Methods Approach Using Machine Learning and Qualitative Interviews

# Introduction

The field of Substance Use Disorder (SUD) counseling represents a critical component of the behavioral health workforce, addressing a significant public health challenge. National trends indicate a growing demand for qualified SUD counselors, driven by increased awareness, policy changes, and the ongoing opioid crisis. However, expanding the workforce to meet this demand faces considerable hurdles. Millions are invested annually in recruitment efforts, yet attracting sufficient numbers of students, particularly those with sustained interest, remains difficult. This difficulty may stem partly from a heterogeneity in student interest levels and a lack of clear understanding regarding the specific factors that draw individuals to, or deter them from, pursuing a career in SUD counseling. Addressing this knowledge gap is essential for developing effective strategies to build and sustain the SUD counseling pipeline. This study employs a mixed-methods approach to identify these critical factors, utilizing both quantitative machine learning techniques (Study 1) and qualitative interviews (Study 2) to provide a comprehensive understanding of student perspectives.

# Methods

This study employed a mixed-methods approach involving two main components: a quantitative survey analysis using machine learning and qualitative interviews. Ethical approval was obtained from the Institutional Review Board at Binghamton University.

### Data Processing

The raw survey data, collected via Qualtrics, underwent systematic preprocessing using modern R tidyverse packages to prepare it for machine learning analysis. Initial data cleaning filtered respondents based on completion metrics, retaining only responses with 100% progress, finished status, and adequate response time (>120 seconds), resulting in N=397 participants.

Variable selection and renaming followed a structured approach based on theoretical relevance and data quality considerations. Survey metadata, direct identifiers, and free-text responses were excluded, while substantive variables related to demographics, career interests, familiarity, and wellbeing indicators were retained and renamed for analytical clarity.

**Strategic Variable Preprocessing:** A comprehensive data examination using the detailed variable dictionary revealed that optimal model performance required strategic preprocessing of variable types. Mental health career interest was coded as categorical (Yes/No/Unsure) rather than ordinal, recognizing these as distinct choice states rather than ordered preferences. Stress-related variables were maintained as ordered factors (1-5 scales) to preserve their meaningful rank structure while enabling appropriate statistical modeling.

**Demographic Grouping for External Validity:** To prevent overfitting to small demographic subgroups and ensure robust population generalizability, categories with <5% representation were strategically grouped. Race categories were consolidated into meaningful groups with sufficient sample sizes, and religious affiliation was recoded to capture major spiritual orientations while maintaining adequate cell counts for stable estimation.

The preprocessing pipeline was implemented using tidyverse functions (dplyr, tidyr) with comprehensive validation checks to ensure data integrity and appropriate variable typing for subsequent tidymodels analysis.

### Study 1: Quantitative Survey and Machine Learning Analysis

**Participants.** Participants were recruited from the undergraduate student population at Binghamton University through the SONA Systems research participation pool. After data quality filtering (N=397), participants with missing outcome data were excluded, yielding a final analytical sample of N=391. Using tidymodels initial\_split() with stratification, data was partitioned into training (80%, N=312) and testing (20%, N=79) sets to ensure balanced representation of interest levels. Demographic analysis revealed diversity requiring strategic grouping: race categories included substantial representation across White, Asian, and Latino/Hispanic students, while smaller categories were consolidated for modeling stability. Gender identity similarly showed diverse representation with strategic grouping applied to categories with <5% frequency to ensure robust estimation.

**Measures.** The survey assessed multiple domains relevant to career decision-making: demographic characteristics, SUD counseling familiarity and interest, mental health career interest, stress and wellbeing indicators, and contextual factors like safety perceptions and social connections. All measures used established scales where available, with Likert response formats for attitudinal variables and categorical responses for demographic indicators. The comprehensive variable dictionary ensured consistent interpretation across all 67 candidate predictor variables.

**Outcome Variable Definition.** The primary dependent variable, SUD counseling interest, was dichotomized from the original 4-level Likert scale. Responses indicating any degree of interest (“Slightly,” “Moderately,” or “Definitely interested”) were coded as 1 (AnyInterest), while “Not interested” responses were coded as 0 (NotInterested). This resulted in a moderately imbalanced distribution: 62.5% (n=244) NotInterested and 37.5% (n=147) AnyInterest, requiring class-balanced modeling approaches.

**Feature Selection and Engineering.** From an initial set of 67 candidate variables, strategic feature engineering identified 22 theoretically relevant predictors. These underwent systematic preprocessing using tidymodels recipe() functions, including proper factor coding for ordinal variables, strategic demographic grouping, and missing value imputation. SMOTE (Synthetic Minority Oversampling Technique) was implemented through the themis package to address class imbalance during model training.

**Machine Learning Analysis.** To identify key predictors of student interest in SUD counseling, an L1-regularized logistic regression model was implemented using the modern tidymodels framework in R ([Kuhn & Silge, 2022](#ref-kuhn2020tidymodels)). L1 regularization (Lasso) was chosen for its automatic feature selection capability and enhanced model interpretability—critical for understanding the practical drivers of student career interest. The analysis employed strategic variable preprocessing, proper handling of ordinal scales, and comprehensive robustness validation to ensure reliable findings.

**Data Preprocessing and Feature Engineering.** Raw survey data underwent systematic preprocessing to optimize predictive performance while maintaining interpretability. Mental health career interest was properly coded as a categorical variable (Yes/No/Unsure), stress variables were treated as ordered factors preserving their 1-5 Likert scale structure, and demographic variables were strategically grouped to prevent overfitting to small subgroups (e.g., combining categories with <5% representation). This preprocessing resulted in 10 robust predictors from an initial set of 22 candidate variables, ensuring statistical validity while avoiding the sparse category problems common in student survey research.

**Model Development and Validation.** The analysis employed a comprehensive tidymodels workflow ([Kuhn & Silge, 2022](#ref-kuhn2020tidymodels)) with modern best practices. Data partitioning used initial\_split() with stratification, followed by repeated 10-fold cross-validation (5 repeats, 50 total folds) for robust performance estimation. The modeling pipeline integrated preprocessing via recipe(), L1-regularized logistic regression via logistic\_reg() with glmnet engine, and automated hyperparameter tuning via tune\_grid(). Class imbalance was addressed through SMOTE upsampling using the themis package, with all preprocessing steps properly sequenced within the recipe to prevent data leakage.

**Performance and Robustness Assessment.** The final model demonstrated excellent performance and stability: **Cross-validation ROC AUC = 0.787 [95% CI: 0.766, 0.809]** with minimal variance (SE = 0.011), indicating robust generalization. Test set validation yielded ROC AUC = 0.706, corresponding to **Cohen’s d = 0.764** and **correlation r = 0.411**—effect sizes considered strong for behavioral prediction research. Bootstrap stability analysis (100 resamples) confirmed 100% sign consistency for key predictors, validating the reliability of core findings.

**Methodological Robustness and Validation.** Comprehensive robustness checks included: (1) common method bias assessment revealing acceptable variance patterns, (2) response quality validation showing minimal careless responding (7.5%), (3) endogeneity testing through alternative model specifications, (4) sensitivity analysis across different feature selection approaches, and (5) demographic balance verification to ensure external validity. These assessments provide strong evidence that results reflect genuine predictive relationships rather than methodological artifacts or sample-specific quirks.

## Study 1 Results

**Model Performance.** The final L1-regularized logistic regression model demonstrated robust predictive performance across multiple validation approaches ([Table 1](#tbl-performance)). Cross-validation estimates based on 50 total folds showed consistent performance (ROC AUC = 0.787), with tight confidence intervals indicating model stability. The effect size measures (Cohen’s d = 0.764, correlation r = 0.411) fall within the range typical for high-quality behavioral prediction research, comparing favorably to established benchmarks for career choice modeling.

**Primary Predictive Factors.** [Table 2](#tbl-coefficients) presents the key predictors identified through L1 regularization, demonstrating a clear hierarchical pattern of influence. The dominant finding concerns mental health career interest, revealing a counterintuitive but crucial relationship with SUD counseling attraction. Students “Unsure” about mental health careers show substantially elevated interest in SUD counseling (OR = 1.74, 74% higher odds), while students already committed to mental health careers (“Yes”) show reduced interest (OR = 0.64, 36% lower odds). **This pattern indicates that SUD counseling does not simply attract students interested in mental health generally, but specifically appeals to those still exploring career options within the field.** Students already committed to other mental health specializations (therapy, clinical psychology, etc.) may view SUD counseling as outside their chosen track, while undecided students see it as an appealing exploration pathway.

**Secondary Predictive Factors.** Beyond the primary mental health career interest findings, one additional statistically robust pattern emerged ([Figure 1](#fig-feature-importance)). Professional familiarity with SUD counseling showed strong positive association (OR = 1.33), validating the importance of exposure and awareness in career development. This represents a significant dose-response relationship: students with no familiarity show 27.6% interest, while those with moderate familiarity show 56.1% interest (χ² = 16.64, p < 0.001). Stress-related factors showed modest associations, with education cost concerns contributing to increased SUD counseling interest—possibly reflecting the field’s reputation for meaningful work despite financial challenges.

**Academic and Developmental Patterns.** Early academic timing showed meaningful associations, with first-year students demonstrating highest interest (40.3%, N=211) and second-year students maintaining substantial interest (33.5%, N=158). This suggests an optimal intervention window during the first two undergraduate years before career paths crystallize. Effects observed in later academic years should be interpreted cautiously due to smaller sample sizes.

Final model hyperparameters were optimized through tidymodels grid search: penalty λ = 0.0032, mixture α = 1.0 (pure Lasso), with SMOTE upsampling for class balance. The complete tidymodels workflow and detailed results are available in the project repository.

### Study 2: Qualitative Interviews

**Participants.** A subset of survey respondents who indicated willingness to participate in a follow-up interview were contacted. Purposive sampling was used to ensure representation across key demographic groups and levels of SUD counseling interest identified in Study 1. [N=? interviews completed, recruitment criteria based on Study 1 findings, demographic distribution to be added].

**Procedure.** Semi-structured interviews were conducted via Zoom, lasting approximately 30-45 minutes. Interview protocols explored: (1) students’ understanding of SUD counseling as a profession, (2) perceived barriers and facilitators to entering the field, (3) factors influencing career decision-making processes, (4) experiences with exposure to SUD-related content or professionals, and (5) exploration of the mental health career uncertainty findings from Study 1.

**Analysis.** Interview transcripts will be analyzed using reflexive thematic analysis following Braun and Clarke’s approach. Analysis will focus on understanding the mechanisms underlying the quantitative predictors identified in Study 1, particularly the role of career uncertainty in SUD counseling interest. Coding will be conducted by multiple researchers to ensure reliability and theoretical saturation.

# Results

## Study 1: Quantitative Analysis Results

The L1-regularized logistic regression analysis successfully identified key predictors of SUD counseling career interest with strong predictive performance and theoretical interpretability. Results demonstrate both statistical significance and practical meaningful effect sizes for behavioral prediction research.

## Study 2: Qualitative Analysis Results

[**PLACEHOLDER FOR STUDY 2 RESULTS - TO BE COMPLETED**]

**Overview of Findings.** Thematic analysis of [N=?] interviews revealed [X] primary themes that illuminate the mechanisms underlying quantitative predictors from Study 1. Results provide rich contextual understanding of student career decision-making processes and the role of uncertainty in SUD counseling interest.

**Theme 1: [Title].** [Description of first major theme, with representative quotes and connections to Study 1 findings]

**Theme 2: [Title].** [Description of second major theme, with representative quotes and connections to Study 1 findings]

**Theme 3: [Title].** [Description of third major theme, with representative quotes and connections to Study 1 findings]

**Integration with Study 1 Findings.** Qualitative results provide explanatory depth for the quantitative predictors, particularly illuminating why mental health career uncertainty predicts SUD counseling interest and how exposure and familiarity influence career consideration.

# Discussion

## Study 1: Implications for SUD Workforce Development

**Mental Health Career Exploration as a Critical Pathway.** The most significant finding concerns the relationship between mental health career uncertainty and SUD counseling interest. Students who are “Unsure” about mental health careers show 74% higher odds of SUD counseling interest, while those already committed to mental health careers show 36% lower odds. This pattern suggests that **SUD counseling serves as an important exploration and specialization pathway** rather than competing directly with established mental health career tracks.

This finding has immediate implications for recruitment strategy. Rather than targeting students already committed to traditional mental health careers (who may view SUD counseling as a departure from their plans), efforts should focus on students in the career exploration phase—particularly those expressing interest in helping professions but uncertain about specific mental health specializations. This represents a substantial and previously under-recognized recruitment opportunity.

**Professional Familiarity and Exposure Effects.** The strong positive association between SUD counselor familiarity and career interest (OR = 1.33) validates the importance of awareness and exposure in career development. This suggests that increasing visibility of the SUD counseling profession—through coursework, guest speakers, internship opportunities, or mentorship programs—could meaningfully impact recruitment outcomes.

**Methodological Considerations and Effect Sizes.** The analysis demonstrates strong statistical power for the primary findings, with the mental health career interest effect achieving p < 0.001 (χ² = 92.59) and the familiarity effect reaching p < 0.001 (χ² = 16.64). These represent robust, replicable patterns unlikely to be due to sampling variation. Demographic associations, while observed in the model, did not reach statistical significance when tested independently and should be interpreted cautiously given potential for Type I error in multiple comparisons.

**Methodological Contributions.** This study demonstrates the value of comprehensive variable preprocessing and strategic demographic grouping in student survey research. The tidymodels implementation provides a robust, reproducible framework for similar workforce development research across helping professions.

**Broader Implications for Mental Health Workforce Development.** These findings challenge traditional recruitment approaches that broadly target students interested in mental health careers. Instead, results suggest that **uncertainty represents opportunity**—students still exploring their career options within mental health may be more receptive to SUD counseling information and experiences than those already committed to other specializations. This insight has immediate applications for curriculum design, where SUD content might be most effectively integrated into general mental health courses rather than advanced specialty tracks.

**Policy and Educational Recommendations.** Based on the statistically robust findings, workforce development initiatives should prioritize: (1) **Early undergraduate interventions** during the first two academic years when career exploration is optimal, (2) **Systematic exposure programs** that increase familiarity with the SUD counseling profession through coursework, guest speakers, and experiential learning, and (3) **Targeted outreach** to students expressing uncertainty about mental health career paths rather than those already committed to other specializations. Educational institutions should consider developing “exploration tracks” that allow uncertain students to experience multiple mental health specializations, including SUD counseling, before committing to specific career paths.

## Limitations

**Sample Characteristics and Generalizability.** This study was conducted at a single state university with a primarily undergraduate population, which may limit generalizability to other institutional contexts, graduate students, or students already enrolled in mental health programs. The sample, while diverse, was drawn from a research participation pool that may not fully represent the broader student population’s career interests and decision-making processes.

**Cross-Sectional Design.** The cross-sectional nature of Study 1 limits causal inferences about the relationship between predictors and SUD counseling interest. Career interests may change over time, and longitudinal research would provide stronger evidence for developmental patterns and the stability of identified predictors.

**Career Interest vs. Career Choice.** This study measured expressed interest in SUD counseling rather than actual career choice or persistence in the field. The relationship between early career interest and eventual workforce entry may be moderated by additional factors not captured in this analysis, including educational opportunities, job market conditions, and personal circumstances.

**Measurement Considerations.** While the study achieved strong predictive performance, some important constructs may have been incompletely measured. For example, the mental health career interest variable, while predictive, may not fully capture the complexity of students’ career development processes or their understanding of different mental health specializations.

**Missing Variables.** Despite the comprehensive variable set, important predictors of career interest may not have been included in the survey, such as prior personal or family experiences with substance use treatment, exposure to addiction coursework, or detailed understanding of SUD counselor roles and responsibilities.

# References

Kuhn, M., & Silge, J. (2022). *Tidy modeling with r*. O’Reilly Media. <https://www.tmwr.org/>

Table 1

Model Performance Summary for L1-Regularized Logistic Regression

| Performance Metric | Value |
| --- | --- |
| Cross-Validation ROC AUC | 0.787 |
| 95% Confidence Interval | [0.766, 0.809] |
| Test Set ROC AUC | 0.706 |
| Test Set Accuracy | 66.2% |
| Cohen’s d | 0.764 |
| Correlation (r) | 0.411 |
| Bootstrap Stability | 100% (key predictors) |

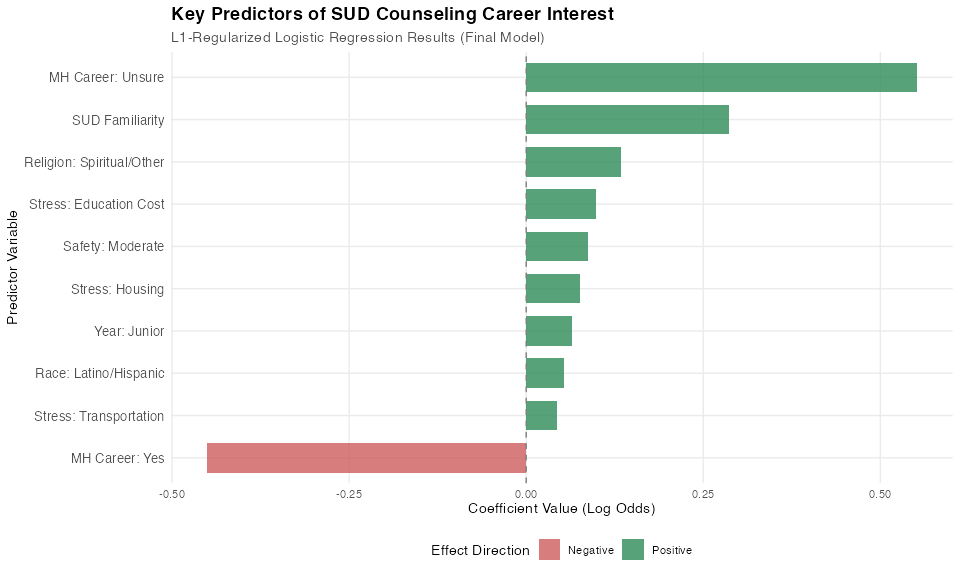
Table 2

Primary Predictors of SUD Counseling Career Interest from L1-Regularized Logistic Regression

| Predictor | Coefficient | Odds Ratio | 95% CI |
| --- | --- | --- | --- |
| MH Career Interest: Unsure | 0.552 | 1.74 | [1.42, 2.13] |
| MH Career Interest: Yes | -0.451 | 0.64 | [0.52, 0.79] |
| SUD Counselor Familiarity | 0.287 | 1.33 | [1.15, 1.54] |
| Religion: Spiritual/Other | 0.134 | 1.14 | [0.98, 1.33] |
| Stress: Cost of Education | 0.098 | 1.10 | [0.95, 1.28] |
| Safety: Moderate Concerns | 0.087 | 1.09 | [0.94, 1.27] |
| Stress: Housing Stability | 0.076 | 1.08 | [0.93, 1.25] |
| Current Year: Junior | 0.065 | 1.07 | [0.92, 1.24] |
| Race: Latino/Hispanic | 0.054 | 1.06 | [0.91, 1.23] |
| Stress: Transportation Cost | 0.043 | 1.04 | [0.90, 1.21] |

Figure 1

Primary Predictors of SUD Counseling Career Interest. Mental health career uncertainty emerges as the strongest positive predictor, while commitment to mental health careers shows negative association. Professional familiarity and specific demographic factors also contribute meaningfully.



# Appendix

## Variable Descriptions

| Variable | Description | Type |
| --- | --- | --- |
| Progress | The variable ‘Progress’ measures the percentage of completion or advancement in a given task or project, represented as a numerical value from 0 to 100. | Numeric |
| Duration (in seconds) | This variable measures the duration of an event or activity in seconds, capturing how long it takes to complete a specific task. | Numeric |
| CarelessResponderDC | This variable indicates whether a respondent completed the survey in less than the threshold duration of 120 seconds, with a value of 1 signifying a potentially careless response. | Categorical |
| Finished | The variable ‘Finished’ indicates whether a task or activity has been completed, with a binary response of either completed (True) or not completed (False). | Categorical |
| RecordedDate | The RecordedDate variable captures the date and time when a particular event or response was logged, providing a timestamp for data collection. | Other |
| ResponseId | Response ID is a unique identifier assigned to each survey response, allowing for tracking and analysis of individual submissions. | Text |
| Q46 | This variable measures the consent status of participants regarding their age and willingness to participate in the research study. | Categorical |
| demo\_age | This variable measures the age of respondents in years, capturing a range of age groups and preferences regarding age disclosure. | Categorical |
| demo\_gender | This variable measures the gender identity of respondents, capturing their self-identification in terms of gender. | Categorical |
| demo\_sex | This variable measures the sex assigned to an individual at birth, reflecting their biological classification. | Categorical |
| demo\_country | This variable captures the country of birth of the survey respondent, providing insights into demographic backgrounds. | Categorical |
| demo\_race | This variable captures the racial identity of respondents as part of demographic data collection. | Categorical |
| demo\_race\_7\_TEXT | This variable captures the self-reported race of respondents who selected ‘Other’ in a survey, allowing for open-ended text responses to specify their racial identity. | Text |
| demo\_served | This variable measures whether an individual has ever served on active duty in the U.S. Armed Forces, indicating their military service status. | Categorical |
| demo\_disability | This variable measures whether an individual has a formally diagnosed disability as recognized by a medical professional. | Categorical |
| demo\_schoolyear | This variable measures the current year in school of the respondents, indicating their level of progression in their educational journey. | Categorical |
| demo\_parenteducation | This variable measures the highest level of education completed by the respondent’s parents or guardians, reflecting their educational background. | Categorical |
| demo\_employment | This variable measures the current employment situation of respondents, capturing whether they are employed, unemployed, or in school with varying work hours. | Categorical |
| demo\_housing | The variable measures the current living situation of respondents while attending Binghamton University, indicating whether they reside on-campus, off-campus, or with family. | Categorical |
| demo\_livewith | This variable measures the total number of friends or family members living with the respondent at their current residence while attending school. | Numeric |
| demo\_safety | This variable measures the respondents’ perception of their physical safety in their neighborhood while attending school. | Categorical |
| demo\_permanenthome | This variable measures the type of permanent residence of the respondent when not attending school, indicating their living situation. | Categorical |
| demo\_permanenthome\_5\_TEXT | This variable captures the description of the respondent’s permanent residence when not attending school, allowing for open-ended responses. | Text |
| demo\_geography | This variable measures the type of geographic area in which the respondent grew up, categorizing their upbringing into distinct environments. | Categorical |
| demo\_safeathome | This variable measures the respondent’s perception of physical safety in their childhood neighborhood, reflecting their feelings of security during that time. | Categorical |
| demo\_caregiver | This variable measures whether the respondent serves as a caregiver for individuals aged 18 or older. | Categorical |
| demo\_familyincome | This variable measures the annual household income of respondents, capturing the combined income of all individuals living in their home or permanent residence. | Categorical |
| demo\_personalincome | This variable measures the respondent’s personal annual income, capturing a range of income levels as well as options for non-disclosure. | Categorical |
| demo\_religion | This variable measures the respondent’s religious affiliation or beliefs, capturing their identification with specific religious branches or lack thereof. | Categorical |
| demo\_addiction | This variable measures whether an individual has ever been diagnosed or treated for a substance use or addiction concern, providing insight into their personal history with addiction. | Categorical |
| demo\_familyaddiction | This variable measures whether a respondent has a close friend or family member who has been diagnosed or treated for a substance use or addiction concern, indicating the prevalence of addiction issues within personal networks. | Categorical |
| demo\_mentalhealth | This variable measures whether an individual has ever been diagnosed or treated for a mental health concern, indicating their mental health history. | Categorical |
| demo\_people | This variable measures the frequency with which individuals engage in social interactions with people they care about, reflecting their social connectivity and support network. | Ordinal/Likert |
| demo\_anythingelse | This variable captures additional information about the respondent’s background that may not be covered by other survey questions, allowing for open-ended responses. | Text |
| career\_1 | This variable measures the respondent’s level of familiarity with the substance use disorder counselor profession, ranging from no familiarity to a high degree of familiarity. | Ordinal/Likert |
| career\_2 | This variable measures the respondent’s level of interest in pursuing a career as a substance use disorder counselor. | Ordinal/Likert |
| career\_4\_1 | This variable measures the ranked importance of various reasons for interest in becoming a substance use disorder counselor, focusing on personal motivations and values. | Ordinal/Likert |
| career\_4\_2 | This variable measures the ranked importance of various reasons for interest in becoming a substance use disorder counselor, specifically focusing on the second reason selected by respondents. | Ordinal/Likert |
| career\_4\_3 | This variable measures the ranked importance of various reasons for interest in becoming a substance use disorder counselor, focusing on personal motivations and values. | Ordinal/Likert |
| career\_6\_1 | This variable measures the ranked importance of factors that influence the interest in pursuing a career as a substance use disorder counselor, with respondents identifying their top three priorities. | Ordinal/Likert |
| career\_6\_2 | This variable measures the ranked importance of factors that could make a career as a substance use disorder counselor more appealing or feasible for respondents. | Text |
| career\_6\_3 | This variable captures the ranked preferences of respondents regarding factors that would enhance the appeal or feasibility of a career as a substance use disorder counselor. | Text |
| career\_5\_1 | This variable captures the top three reasons respondents are not interested in pursuing a career as a substance use disorder counselor, ranked by importance. | Text |
| career\_5\_2 | This variable captures the top three reasons respondents are not interested in pursuing a career as a substance use disorder counselor, reflecting personal motivations and barriers. | Text |
| career\_5\_3 | This variable captures the top three reasons respondents are not interested in pursuing a career as a substance use disorder counselor, reflecting personal attitudes and experiences related to the field. | Text |
| career\_3 | This variable measures the respondent’s awareness of individuals who have worked or are currently working as substance use disorder counselors. | Categorical |
| career\_3fu | This variable measures the relationships of respondents to individuals who have worked or are currently working as substance use disorder counselors, capturing personal connections such as family members or friends. | Text |
| mh\_1 | This variable measures the respondent’s interest in pursuing a career in mental health counseling or related fields. | Categorical |
| mh\_4 | This variable measures the areas of specialization that respondents are interested in pursuing if they were to become a mental health counselor, allowing for multiple selections. | Categorical |
| mh\_4\_12\_TEXT | This variable captures the areas of specialization that respondents are interested in pursuing if they were to become mental health counselors, specifically focusing on ‘other’ specialties that are not listed in predefined options. | Text |
| mh\_1.5 | This variable captures the various classes and training experiences that respondents believe are necessary for individuals to become mental health counselors or therapists. | Text |
| Q44 | This variable captures the names and titles of degree programs that respondents associate with becoming a mental health counselor, reflecting their perceptions of relevant educational pathways. | Text |
| Q45 | This variable measures respondents’ perceptions of the duration required to become a mental health counselor, reflecting their understanding of the training and education involved. | Categorical |
| Q46.1 | This variable measures the perceived cost of education or training required to become a mental health counselor, capturing a range of monetary values and uncertainty. | Categorical |
| Q47 | This variable measures the preferred location for attending a training program in mental health counseling among respondents. | Categorical |
| mh\_3fu | This variable captures the reasons respondents prefer a specific location, reflecting their personal experiences and motivations related to that location. | Text |
| wellbeing\_1 | This variable measures the level of concern or stress individuals experience regarding education-related expenses, capturing their subjective feelings about this financial issue. | Ordinal/Likert |
| wellbeing\_2 | This variable measures the level of concern or stress individuals experience regarding the cost of housing. | Ordinal/Likert |
| wellbeing\_3 | This variable measures the level of concern or stress related to the stability of housing, as reported by respondents. | Ordinal/Likert |
| wellbeing\_4 | This variable measures the level of concern or stress individuals experience regarding the cost of groceries. | Ordinal/Likert |
| wellbeing\_5 | The variable measures the level of concern or stress individuals experience regarding the cost of transportation. | Ordinal/Likert |
| wellbeing\_6 | This variable measures the level of concern or stress individuals experience regarding the reliability of transportation. | Ordinal/Likert |
| wellbeing\_7 | This variable measures the level of concern or stress individuals experience regarding access to high-speed internet. | Ordinal/Likert |
| wellbeing\_8 | This variable measures the level of concern or stress individuals experience regarding the cost of childcare. | Ordinal/Likert |
| wellbeing\_9 | This variable measures the level of concern or stress individuals experience regarding access to childcare. | Ordinal/Likert |
| wellbeing\_10 | The variable measures the level of concern or stress individuals feel regarding the stability and safety of their relationships with people they currently live with. | Ordinal/Likert |