

## **Appendix: Construction of Predictor Variables**

- Gender (Binary): This variable represents the gender of each respondent. The variable was constructed from question **a4** (What is your gender?) of the Core Module. Raw data was dichotomously coded (0 = Male; 1 = Female).
- Gender Minority Status (Binary): This variable represents the gender minority status of each respondent. Gender minority status describes individuals whose gender identity is different from the sex they were assigned at birth. The variable was constructed from question **a5** (Some people describe themselves as transgender when how they think or feel about their gender is different from the sex they were assigned at birth. Are you transgender?) of the Core Module. Responses were categorically coded (1 - No, I am not transgender; 2 - Yes, I am transgender; 3 - I am not sure if I am transgender; 4 - Decline to respond). Raw data was recoded to create a binary variable, with raw data value 1 being recoded as 0 = cisgender (No), and the raw data values 2 - 4 were recoded as 1 = transgender (Yes).
- Sexual Minority Status (Binary): This variable represents the sexual minority status of each respondent. Sexual minority status described individuals who identify as lesbian, gay, bisexual, queer, or additional non-heterosexual identity. The variable was constructed from question **a18** of the Core Module. Responses were categorically coded (1 - Straight [not gay]; 2 - Gay or Lesbian; 3 – Bisexual; 4 - I am not sure yet; 5 - Something else; 6 - Decline to respond). Raw data was recoded to create a binary

variable, with raw data value of 1 being recoded as 0 = heterosexual (No), and the raw data values 2 - 6 were recoded as 1+ non-heterosexual (Yes).

- Academic Performance (Continuous): This variable represents the degree of academic performance among each respondent. The variable was constructed from the **a20** question (During the past 12 months, how would you describe the grades you mostly received in school?) from the Core Module. This variable was then standardized so the variable has a mean of zero and a standard deviation of one. Scores ranged from -2.54 to 1.30 with a high score signifying a higher degree of academic performance.
- Lifetime Use of Alcohol (Continuous): This variable represents the frequency a respondent consumes alcohol. The variable was constructed from the **a51\_a52** question from the substance use component of the Core Module designed to assess the extent to which participants consumes mind altering substances. The raw data was then standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -0.52 to 2.46 with a higher score signifying a higher frequency of substance use.
- Lifetime Use of Cigarettes (Continuous): This variable represents the frequency a respondent engages in smoking cigarettes. The variable was constructed from the **a48\_a49** question from the substance use component of the Core Module. The raw data was then standardized so the variable had a mean of zero and a standard deviation of one.

Scores ranged from -0.23 to 5.31 with a higher score signifying a higher frequency of substance use.

- Lifetime Use of Vapes (Continuous): This variable represents the frequency a respondent engages in smoking via a vaporizer or vape product. The variable was constructed from the **a50\_a51** question from the substance use component of the Core Module. The raw data was then standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -0.46 to 2.70 with a higher score signifying a higher frequency of substance use.
- Lifetime Use of Cannabis (Continuous): This variable represents the frequency a respondent consumes cannabis. The variable was constructed from the **a52\_a53** question from the substance use component of the Core Module. The raw data was then standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -0.49 to 2.42 with a higher score signifying a higher frequency of substance use.
- Lifetime Use of other Drugs (Continuous): This variable represents the frequency a respondent consumes substances not previously mentioned. The variable was constructed by merging responses from 11 questions from the substance use component of the Core Module. A continuous variable was constructed by summing the responses to the 11 questions. The raw data was then divided by the number of responses and standardized so

the variable had a mean of zero and a standard deviation of one. Scores ranged from -0.26 to 6.47 with a higher score signifying a higher frequency of substance use.

- Foster Care Placement (Binary): This variable represents if a respondent is currently in a foster care placement. The variable was constructed from question **a9** (What best describes where you live?) of the Core Module. Responses were categorically coded (1- A home with one or more parents/guardian; 2 - Other relative's home; 3- A home with more than one family; 4 - Friend's home; 5 - Foster home, group care, or waiting place; 6 - Hotel or motel; 7 - Shelter, car, campground, or other temporary housing; 8 - Other living arrangement). Raw data was recoded to create a binary variable, with raw data values 1 – 3, and 5 – 8 being recoded as 0 – No, and the raw data value of 4 recoded as 1- Yes.
- Parent Education Level (Continuous): This variable represents the degree of education for each respondents' parents or guardians. The variable was constructed from the **a10** question (What is the highest level of education your parents or guardians completed?) from the Core Module. The raw data was then standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -2.54 to 1.30 with a higher score signifying a higher level of parental education.
- Positive Substance Use Education (Continuous): This variable represents the degree a respondent has received and internalized positive substance use education. The variable was constructed by merging responses from eight questions from the Core Module. A

continuous variable was constructed by summing the responses to the eight questions.

The raw data was then divided by the number of responses and standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -1.81 to 1.03 with a higher score signifying a higher level of internalized education.

- School-Based Victimization (Continuous): This variable represents the frequency a respondent has experienced instances of victimization including bullying, physical or sexual violence, or other processes of where an individual is harmed. The variable was constructed by merging responses from 24 questions from the Core Module. A continuous variable was constructed by summing the responses to the 24 questions. The raw data was then divided by the number of responses and standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -0.64 to 4.48 with a higher score signifying a higher frequency of experiences of victimization.
- Access to Drugs and Alcohol (Continuous): This variable represents the degree of ease of access a respondent has to obtain drugs and/or alcohol. The variable was constructed by merging responses from four questions from the Core Module. A continuous variable was constructed by summing the responses to the four questions. The raw data was then divided by the number of responses and standardized so the variable had a mean of zero and a standard deviation of one. Scores ranged from -2.24 to 1.00 with a higher score signifying an easier degree of access to obtaining drugs and/or alcohol.

- Depressive Symptoms (Binary): This variable represents the presence of depressive systems for each respondent. The variable was constructed from question **a130\_a118** (During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more that you stopped doing some usual activities?) of the Core Module. Raw data was dichotomously coded (0 = No; 1 = Yes).
- Homelessness (Binary): This variable represents if a respondent consumed breakfast the day the survey was administer. Responses were categorically coded (1- A home with one or more parents/guardian; 2 - Other relative's home; 3- A home with more than one family; 4 - Friend's home; 5 - Foster home, group care, or waiting place; 6 - Hotel or motel; 7 - Shelter, car, campground, or other temporary housing; 8 - Other living arrangement). Raw data was recoded to create a binary variable, with raw data values 1 – 5 recoded as 0 = No, and the raw data values of 6 - 8 recoded as 1 = Yes.