**EDUR 9131: Reliability and Reverse Coding**

**18 February 2023**

**1. Menon’s Questionnaire**

Collect data, unique numeric ID needed.

**2. Data Description**

* Data collected from 500+ undergraduate students at Georgia Southern
* Study focus was on cyber-harassment
* Analyses that follow will focus on the following variables: student sex, academic functioning, and life functioning
* Academic Functioning consists of
  + GPA
  + Graduation Confidence
  + University Connectedness
  + Academic Control
  + Intrinsic Motivation
  + Identified Regulation
* Life Functioning consists of
  + Toxic Disinhibition
  + Impulsivity
  + Empathy
  + Stress
  + Life Satisfaction
  + Socially Connected
* Counts of victimization and perpetration among respondents: VictimCount, BullyCount

Student Sex

Shape, rectangle

Description automatically generated

GPA

Text, email

Description automatically generated

Graduation Confidence

Graphical user interface

Description automatically generated with medium confidence

Graphical user interface, application, Word

Description automatically generated

University Connectedness

Table

Description automatically generated

Intrinsic Motivation and Identified Regulation

Table

Description automatically generated

Academic Control

Calendar

Description automatically generated

Empathy

Table

Description automatically generated

Stress

Graphical user interface, application, table

Description automatically generated

Toxic Disinhibition

A picture containing graphical user interface

Description automatically generated

Impulsivity

Calendar

Description automatically generated

Life Satisfaction and Socially Connected

A picture containing table

Description automatically generated

Victim Count and Bully Count (spoken harassment example, also includes written, visual, hacking/impersonating, and social harassment)

Text

Description automatically generated

**3. Reliability: Test-retest, Parallel Form, and Single-Item Reliability (test-retest)**

* Consistency vs Agreement – Used to assess whether two or more ratings, or scores from raters, are similar.
* Consistency: sets of scores show similar rankings or relative positions, but the two sets may have large numeric differences.
* Agreement: sets of scores show both similar rankings/relative positions and have numeric similarities
* Used for test-retest and parallel forms reliability assessment. Used when one may collect scores on multiple occasions **from the same participants**.
* Pearson r can be used to assess consistency, but not agreement because it cannot assessment numeric equivalence.
* ICC = Intraclass correlation coefficient must be used for assessing agreement and can also be used to assess consistency
* Used for quantitative data (ordinal with multiple responses, interval, ratio data)

Table

Description automatically generated

Analysis

* Pearson r to confirm .91 level of consistency between tests 1 and 2 in the first example
* Use ICC
  + SPSS Scale,
  + Reliability analysis,
  + Statistics,
  + Intraclass correlation coefficient,
  + Two-way mixed
  + Absolute Agreement
  + Result – focus on Single Measures ICC

We will use ICC again for assessing rater agreement for ordinal, interval, and ratio data.

**4. Menon’s Questionnaire**

Re-collect data; unique numeric ID needed. Single-item assessment.

**5. Reliability: Internal Consistency**

A picture containing diagram

Description automatically generated

* Do items designed to measure a latent variable show consistency in responses? If yes, that means the scores are internally consistent.
* Cronbach’s alpha most common measure of internal consistency
  + 0.90+ excellent consistency
  + 0.80 very good
  + 0.70 usable for research purposes, typically minimum level needed for research
  + 0.60 usable but may behave poorly, may reduce statistical power to detect differences or relationship.
* Alpha = proportion of shared variance among items, higher alpha higher variance shared, or higher common variance
* Item Analysis, part of Cronbach’s alpha
  + Helps determine item fit among all items used to measure a construct
  + Examine
    - correlations among items
    - negative correlations signal possible reversed items
    - weak correlations indicate items that do not function well
    - corrected item-total correlation: item x with all other items combined
    - Alpha if item deleted: how alpha changes if an item is deleted
  + Don’t let item decisions be mechanical, consider theoretical importance and contribution of item to construct before removing it should item fit be poor

University connectedness (items 35 36 37 38 39)

What is alpha?

Review item fit and contribution.

Toxic Disinhibition (items 16 17 18 19)

What is alpha?

Review item fit and contribution.

Graduation Confidence (items 4 5 6 7)

What is alpha?

Review item fit and contribution.

**6. Reverse Scoring of Items**

* Needed to reverse those item that have different polar or scale response orientations
* Use logic to identify reversed items: assume extreme position, answer all items, and it should be obvious which items cause reposes to flip from high to low scores,
* and then items identified as reversed can be verify via correlations
* formula to reverse

**Reversed Score = (minimum score) + (maximum score) – actual score**

* Can check reversed scores by correlating reversed item with original item
* Also check frequencies for both original and reversed items

Graduation Confidence (items 4 5 6 7)

Identify and reverse items then recheck alpha using reversed items in place of original items

Review item fit and contribution.

Academic control (items 40 41 42 43 44 45 46 47)

Identify and reverse items then recheck alpha using reversed items in place of original items

Review item fit and contribution.

Socially connected (items 13 14 15)

Unique, odd situation with this construct: what do high scores mean?

Identify and reverse items then recheck alpha using reversed items in place of original items

Review item fit and contribution.

**7. Composite Scores**

* Composite scores are constructed scores
  + Summing responses across items or indicators (not a good option, explained below)
  + Mean of responses across items or indicators (good option)
  + Weighted composite from factor analysis or similar analysis (usually sample dependent) is sometimes used; weighted means some items account for more of the composite score than others; this requires more complex statistics or theoretical guidance; using equally weighted composite scores – like taking the sum of all items or the mean of all items – works well in many cases. Weights described below.
* Sometimes called scale scores, but this can be confusing since scale scores generally are understood to be scores with predefined mean and standard deviation (standard score, Z score)
* Sum scored can be misleading
  + Example from test anxiety, student has maximum anxiety
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = 7 (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  + Minimum and maximum summed scores are
    - 1+1+1 = 3
    - 7+7+7 = 21
  + Respondent’s sum = 7+7+7 = 21, which is top score possible for sum of these three items
  + Item 2 has missing data
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = missing (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  + Sum = 7+missing+7 = 14, which is toward middle range of 3 to 21, so misleading level of anxiety presented
* Mean score can ignore missing responses so it reflects better level of anxiety, also mean lies within original scale units so easier to interpret
  + Example from test anxiety, student has maximum anxiety
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = 7 (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  + Scale ranges from 1 to 7
  + Respondent’s mean = 7+7+7 = 21/3 = 7.0 top score possible
  + Item 2 has missing data
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = missing (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  + Mean of available data = 7+7 = **14 / 2** = **7.0**, top score possible
  + Mean of all items = 7+7 = **14 / 3** = **4.66**, misleading score

Be sure calculation of mean uses available data, not all possible scores

Form composite scores for each below and find the correlations among these constructs (use composite scores, not individual items, to assess correlations among constructs)

Be sure to use the appropriate items – reversed scored items where needed..

University connectedness (items 35 36 37 38 39)

Toxic Disinhibition (items 16 17 18 19)

Graduation Confidence (items 4 5 6 7)