Harm Reduction Development and Validation

Zach Budesa

Introduction

This file documents the ongoing process of analysis and visualization of the Harm Reduction Development and Validation Studies. All available study materials, data, code, preregistrations, and final scale can be accessed via OSF: https://osf.io/ejcyt/.

Item Generation

Item generation was undertaken in several stages. The first step in item generation consisted of items generated in two scales (harm reduction principles and harm reduction strategies) which were drafted by the authors and piloted with harm reduction workers in local training and harm reduction supply distribution evaluation. Following the initial pilot project, we revised and added items to address gaps that had been noted. The items were generated with regard to both the SAMHSA Harm Reduction Framework and the National Harm Reduction Coalition's Harm Reduction Principles(https://harmreduction.org/about-us/principles-of-harm-reduction/). Once the initial item pool was generated, we consulted with Dr. David Frank, a harm reduction expert and researcher. Based on his consultation, items were further modified to be accessible to a general population.

We used Wolf et al's (2023) Response Process Evaluation (RPE) method to ensure that a non-expert, non-harm reduction audience would understand the item content. Over X rounds of data collection, we used Prolific to collect qualitative responses to each item. As a group, we evaluated each set of responses to determine whether respondents understood or did not understand the intended meaning of each item. Based on respondent answers and suggestions, we identified items which items were generally understood, and which items should be modified or removed. The final item pool contained 45 items related to harm reduction strategies and principles. We phrased items as simply and directly as possible, occasionally splitting complex topics (experiences of discrimination, other topic) across multiple items when necessary. For study 1, items were answered on a 6-point ordinal scale from 1 (Strongly Disagree) to 6 (Strongly Agree). Following inspection of item responses, studies 2 and 3 used 5-point ordinal scales from 1 (Strongly Disagree) to 5 (Strongly Agree). Participants answered items in a random order, and completed a brief demographic survey.

Items

```r
library(gt)
library(tidyverse)

items <- readr::read_csv("references/codebooks/20231011_hr-scale-codebook.csv")

items %>%
  filter(
    grepl("q", var_label)) %>%
  select(-1) %>%
  gt() %>%
  tab_header(
    title = "Table 1: Initial Item Set"
  ) %>%
  cols_label(
    var_label = "Item Label",
    item = "Item Wording"
  )
```

Table 1: Initial Item Set

Item Label	Item Wording
q1	People who inject drugs should be able to do so in a way that prevents them from causing further
q2	People should have access to tools for safer sex (like condoms, STD tests).
q3	People who use drugs should have access to naloxone/NARCAN.
q4	The general public should have access to naloxone/NARCAN.
q5	Police officers should have access to naloxone/NARCAN.
q6	People who use drugs should have access to tools to test what's in their drugs.
q7	People who use drugs should have access to safe injection supplies (sterile needles and syringes).
q8	People who use drugs should have access to safe inhalation supplies (glass stems and pipes).
q9	People who actively use drugs should have access to therapy/counseling.
q10	People who return to using drugs after a period of abstinence should be allowed to continue in tre
q11	Sobriety should be required for treatment.
q12	Medications used to treat addiction (buprenorphine, naltrexone, or methadone) are an appropriate
q13	Possession of "drug paraphernalia", like syringes and pipes, should be legal.
q14	People who seek medical assistance for overdoses should be protected from drug charges, arrests, a
q15	Sobriety should not be a requirement to access public housing.
q16	Possession of all drugs should be decriminalized (possession would not lead to legal repercussions).
q17	It should be legal for adults to purchase drugs from a dispensary/shop.
q18	People who use drugs should have access to supervised places where they can consume drugs safel
q19	People who use drugs should have access to a legal, non-contaminated drug supply.
q20	People use drugs to escape.
q21	People should be able to use drugs safely.
q22	People who use drugs should be treated with respect.
q23	Poverty affects the health of people who use drugs.
q24	Racism affects the health of people who use drugs.
q25	Gender-based discrimination affects the health of people who use drugs.
q26	Some ways of using drugs are safer than others.
q27	People who use drugs deserve to live good lives.
q28	Reducing drug use is a reasonable goal for people who use drugs.
q29	Some people who use drugs cannot be expected to quit immediately.
q30	People who use drugs should be involved in creating the programs and policies that serve them.
q31	People in recovery from drug use should be involved in creating the programs and policies that ser
q32	Relapse may be a part of the recovery process.
q33	It is possible to live a healthy life without stopping drug use.
q34	People who use drugs should be forced into treatment.
q35	Using drugs is immoral.
q36	Drug use has benefits.
q37	Harm reduction complements traditional addiction prevention, treatment, and recovery services.

q38	People who use drugs benefit society.
q39	Reducing the negative consequences of drug use encourages more people to use drugs.
q40	People will use more drugs if it is safer.
q41	People who use drugs will naturally end up homeless.
q42	Drugs make the world worse.
q43	Drug use will always be part of society.
q44	Chaotic drug use is a rational response to experiences like trauma, homelessness, hunger, and pove
q45	People who use drugs should be able to use medications used to treat addiction (buprenorphine, n

Study 1

Study 1 was preregistered, which can be accessed at https://osf.io/yt7sm. In this study, we conducted exploratory factor analysis using an oblimin rotation based on a polychoric correlation matrix using maximum likelihood estimation for factor extraction. We used the 40-30-20 rule to determine factor loading cut off, which means that retained items should load on their primary factor above 0.4, load onto alternative factors below 0.3, and show a difference of 0.2 between primary and alternative factor loadings. Finally, scale reliability will be assessed using Cronbachs' Alpha, the Greatest Lower Bound, and average split half reliability.

Participants

```
# Libraries
library(psych)
library(correlation)
library(EFA.dimensions)
library(GPArotation)

# Import Data
df <- read.csv("data/clean/20231011_hr-scale-exploratory-data.csv")

# Build Demographic Data Set
demos <- df %>%
  select(age, gid1, race, ed, sud_hx)
```

```
# Build Data Set of Items
hr <- df %>% # N = 301
  select(starts_with("q"), -Q_RecaptchaScore)
```

Through Prolific, we recruited 301 participants for Study 1. These participants were 50% men (n = 149), 70% White (n = 210), 41% had completed a bachelor's degree (n = 122), 88% reported no history of a substance use disorder (SUD) diagnosis, and had an average age of 36. Table 2 provides additional detail about this sample.

```
library(gtsummary)

demos %>%
  tbl_summary(
    missing = "no",
    statistic = list(age = "{mean} ({sd})"),
    digits = list(age ~ c(2,2)),
    sort = list(everything() ~ "frequency"),
    label = list(
      age ~ "Age", gid1 ~ "Gender Identity",
      race ~ "Race/Ethnicity", ed ~ "Highest Level of Education",
      sud_hx ~ "History of Substance Use Disorder Diagnosis")
              ) %>%
  as_gt() %>%
  tab_header(title = "Table 2: Study 1 Participants")
```

Table 2: Study 1 Participants

Characteristic	**N = 301**[1]
Age	35.92 (11.28)
Gender Identity	
Man	149 (50%)
Woman	143 (48%)
Non-Binary or Gender Fluid	7 (2.3%)
Prefer not to say	1 (0.3%)
Race/Ethnicity	
White	210 (70%)
Black or African American	30 (10.0%)

Asian or Asian American	28 (9.3%)
Multiracial	13 (4.3%)
Latine or Hispanic	12 (4.0%)
Prefer not to say	4 (1.3%)
Indigenous, Aboriginal, or First Nations	2 (0.7%)
Prefer to self-describe	2 (0.7%)
Highest Level of Education	
Bachelor's Degree	122 (41%)
High School	97 (32%)
Master's Degree	39 (13%)
Associate's Degree	28 (9.3%)
Doctorate Degree (PhD, PsyD, etc.)	4 (1.3%)
Professional Degree (J.D., M.D., etc.)	4 (1.3%)
Less than High School	3 (1.0%)
Other option not represented here	2 (0.7%)
Prefer not to say	1 (0.3%)
History of Substance Use Disorder Diagnosis	
No	265 (88%)
Yes	28 (9.3%)
Prefer not to say	5 (1.7%)
I don't know	3 (1.0%)

[1] Mean (SD); n (%)

Initial Data Inspection

```
# Create Polychoric Correlation Matrix
hrMatrix <- POLYCHORIC_R(hr, verbose = FALSE)

bart <- cortest.bartlett(hrMatrix, n = 301)

kmo <- KMO(hrMatrix)
```

Using a polychoric correlation matrix, we analyzed Bartlett's Sphericity and the Kaiser-Meyer-Olkin test to ensure that the data were appropriate for factor analysis. Our data appeared adequate for factor analysis, with Kaiser Meyer-Olkin = 0.9189898, and Barlett's sphericity $\chi^2(990) = 1.1486222 \times 10^4$, p = 0. Inspection of the correlation matrix showed many inter-item correlations > .40, indicating that a factor solution should be possible.

Local Dependence Between Items

```
# Saved because LOCALDEP takes forever to run
# ld <- LOCALDEP(hr)
# saveRDS(ld, "study1-local-dependence.rds")

ld <- readRDS("study1-local-dependence.rds")
```

Initial inspection did reveal excessively high inter-item correlations that caused concern. First, there were 14 inter-item correlations which were > .70, and 2 partial correlations > .60. Using the EFA.dimensions package, inspection of possible local dependence was done via Q3, X2, and and G2 statistics. Possible locally dependent items were inspected pair-by-pair. Those that were highly similar were removed, while those that appeared statistically locally dependent but different content wise were retained. Table 3 provides additional detail about possible local dependence. Based on both quantitative and qualitative inspection of items, the following items were removed due to their similarity with others: Items 4 and 5 (local dependence with item 3), item 8 (local dependence with item 7), item 45 (local dependence with item 12), item 22 (local dependence with item 27), items 24 and 25 (local dependence with item 23), item 32 (local dependence with item 29), item 31 (local dependence with item 30), and item 40 (local dependence with item 39). While there are other possible removals, these items represented the most overlap in both statistical and content terms.

```
ld$localdep_stats %>%
  # Remove additional statistic
  select(-JSI) %>%
  # Filter out unconcerning items
  filter(
    abs(Q3) > .34 &
     ( p_X2 < .05 |
```

```r
      p_G2 < .05)
  ) %>%
  gt(groupname_col = "Item_A",
     rowname_col = "Item_B") %>%
  tab_spanner(label = "Correlations", columns = c(3,4)) %>%
  tab_spanner(label = "X2 Statistics", columns = c(6,7)) %>%
  tab_spanner(label = "G2 Statistics", columns = c(8,9)) %>%
  tab_header(title = "Table 3: Inspection of Local Dependence between Similar Items")
```

Table 3: Inspection of Local Dependence between Similar Items

	Correlations			X2 Statistics		G2 Statistics	
	r	partial_r	Q3	X2	p_X2	G2	p_G2
q3							
q4	0.829	0.705	0.688	117.534	0.000	131.412	0.000
q5	0.664	0.531	0.517	84.571	0.000	94.799	0.000
q4							
q5	0.692	0.576	0.570	96.480	0.000	109.867	0.000
q7							
q8	0.852	0.546	0.463	54.659	0.001	70.158	0.000
q12							
q45	0.576	0.392	0.391	51.460	0.001	56.087	0.000
q22							
q27	0.658	0.409	0.382	49.280	0.003	55.253	0.000
q23							
q24	0.497	0.417	0.421	83.618	0.000	79.173	0.000
q24							
q25	0.706	0.632	0.631	160.541	0.000	149.798	0.000
q29							
q32	0.505	0.365	0.367	72.966	0.000	50.120	0.002
q30							

q31	0.732	0.582	0.575	148.996	0.000	152.797	0.000
q33							
q36	0.631	0.488	0.489	101.703	0.000	104.794	0.000
q38	0.614	0.440	0.443	88.980	0.000	96.442	0.000
q42	-0.582	-0.417	-0.424	68.425	0.000	71.765	0.000
q35							
q41	0.615	0.362	0.347	55.910	0.000	62.287	0.000
q42	0.582	0.338	0.347	43.785	0.011	53.620	0.001
q36							
q38	0.628	0.453	0.456	64.551	0.000	75.351	0.000
q42	-0.648	-0.504	-0.510	88.178	0.000	91.166	0.000
q38							
q42	-0.660	-0.497	-0.501	87.129	0.000	96.960	0.000
q39							
q40	0.694	0.587	0.583	104.985	0.000	113.625	0.000

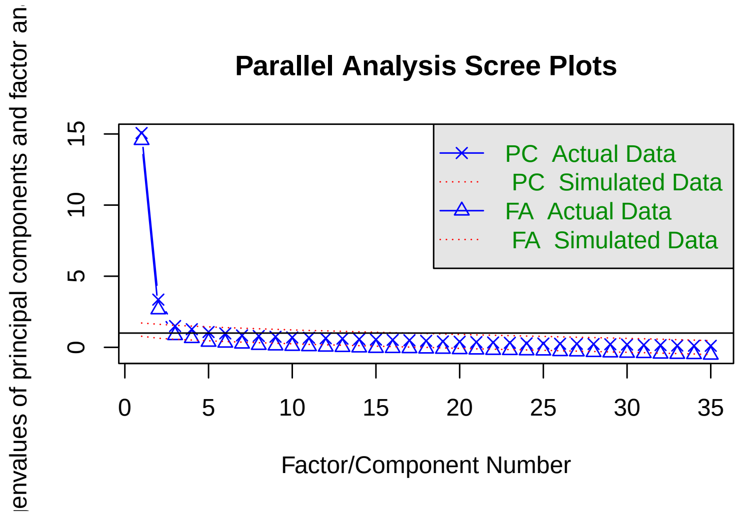
```r
hrMatrix <- POLYCHORIC_R(hr %>%
                              select(-c(q4:q5, q8, q45, q24, q25,
                                        q32, q22, q40, q31)),
                         verbose = FALSE)

bart <- cortest.bartlett(hrMatrix, n = 301)

kmo <- KMO(hrMatrix)
```

After removing items which showed excessive local dependence, bartlett's sphericity and Kaiser-Meyer-Olkin were checked again, with Kaiser Meyer-Olkin = 0.9365298, and Barlett's sphericity $\chi^2(595) = 7566.2975942$, p = 0.

Exploratory Factor Analysis

Parallel Analysis

```
fa.para <- fa.parallel(hrMatrix, n.obs = 301,
                       fm = "ml",
                       cor = "poly")
```




```
Parallel analysis suggests that the number of factors =  4  and the number of components =  2
```

Model 1

```
# Initial Factor Analysis
efa <- EFA.dimensions::EFA(hrMatrix,
                    extraction = "ml",
                    Nfactors = 4,
                    Ncases = 301,
                    rotation = "oblimin",
                    verbose = FALSE)



initial <- efa$pattern %>%
```

```r
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_style(
    locations = list(
      # Factor 1
      cells_body(
      columns = `Factor 1`,
      rows = abs(`Factor 1`) < .30 ),
      # Factor 2
      cells_body(
        columns = `Factor 2`,
        rows = abs(`Factor 2`) < .30
      ),
      # Factor 3
       cells_body(
        columns = `Factor 3`,
        rows = abs(`Factor 3`) < .30
      ),
      # Factor 4
       cells_body(
        columns = `Factor 4`,
        rows = abs(`Factor 4`) < .30
      )),
    style = list(cell_text(color = 'gray'))) %>%
  fmt_number(
    columns = starts_with("Factor"),
    decimals = 3)
```

An initial factor analysis revealed that a four factor solution provided poor fit for the data, with factor 1 explaining 41% of the total variance (eigenvalue = 14.57), factor 2 explaining 8% of the total variance (eigenvalue = 2.8), factor 3 explaining 3% of the total variance (eigenvalue = 1.05), and factor 4 explaining 2.5% of the total variance (eigenvalue = .90). This model had the following model fit indices: $\chi^2(461) = 1130.61$, p = $4.5327627 \times 10^{-58}$; CLI = 0.904; TLI = 0.876; RMSEA = 0.07. In addition to poor fit, there were numerous cross-loadings which exceeded the limit of difference of .2. Table 4 provides full factor loadings.

```r
# New Matrix
hrMatrix <-
  POLYCHORIC_R(
    hr %>%
      select(
        # Factor 1
        q1, q3, q6, q7, q13, q16, q18, q19, q21,q39,
        # Factor 2
        q33, q36, q38, q41, q42,
        # Factor 3
        q2, q9, q10, q23, q27, q29, q30
        ),
    verbose = FALSE)


efa.final <- EFA.dimensions::EFA(hrMatrix,
                    extraction = "ml",
                    Nfactors = 3,
                    Ncases = 301,
                    rotation = "oblimin",
                    verbose = FALSE)

final <- efa.final$pattern %>%
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_style(
    locations = list(
      # Factor 1
      cells_body(
      columns = `Factor 1`,
      rows = abs(`Factor 1`) < .30),
      # Factor 2
      cells_body(
        columns = `Factor 2`,
        rows = abs(`Factor 2`) < .30
      ),
      # Factor 3
       cells_body(
        columns = `Factor 3`,
        rows = abs(`Factor 3`) < .30
```

```
      )),
    style = list(cell_text(color = 'gray'))) %>%
  fmt_number(
    columns = starts_with("Factor"),
    decimals = 3)


left_join(
    initial$`_data`,
    final$`_data`,
    by = c(
      "Item" = "Item"
    )) %>%
  arrange(
    factor(Item,
           levels =
             c(
               # Factor 1
               "q1", "q3", "q6", "q7", "q13", "q16", "q18", "q19", "q21","q39",
               # Factor 2
               "q33", "q36", "q38", "q41", "q42",
               # Factor 3
               "q2", "q9", "q10", "q23", "q27", "q29", "q30"))) %>%
  janitor::clean_names() %>%
  gt() %>%
  tab_style(
    locations = list(
      # Factor 1
      cells_body(
      columns = factor_1_x,
      rows = abs(factor_1_x) < .30),
      # Factor 2
      cells_body(
        columns = factor_2_x,
        rows = abs(factor_2_x) < .30
      ),
      # Factor 3
       cells_body(
        columns = factor_3_x,
```

```
      rows = abs(factor_3_x) < .30
    ),
    # Factor 4
     cells_body(
      columns = factor_4,
      rows = abs(factor_4) < .30
    ),
     # Factor 1 - Final
    cells_body(
    columns = factor_1_y,
    rows = abs(factor_1_y) < .30),
    # Factor 2 - Final
    cells_body(
      columns = factor_2_y,
      rows = abs(factor_2_y) < .30
    ),
    # Factor 3 - Final
     cells_body(
      columns = factor_3_y,
      rows = abs(factor_3_y) < .30
    )),
  style = list(cell_text(color = 'gray'))) %>%
fmt_number(
  columns = starts_with("Factor"),
  decimals = 3)  %>%
fmt_missing() %>%
tab_header(
  title = "Table 4: Initial and Final Factor Structure"
) %>%
tab_spanner(
  label = "Initial 4 Factor Solution",
  columns = 2:5
) %>%
tab_spanner(
  label = "Final 3 Factor Solution",
  columns = 6:8
) %>%
cols_label(
  factor_1_x  = "Factor 1",
```

```
    factor_2_x = "Factor 2",
    factor_3_x = "Factor 3",
    factor_4 = "Factor 4",
    factor_1_y = "Factor 1",
    factor_2_y = "Factor 2",
    factor_3_y = "Factor 3"
  )
```

Table 4: Initial and Final Factor Structure

	Initial 4 Factor Solution				Final 3 Factor Solution		
item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3
q1	0.958	0.089	0.019	−0.032	0.983	0.107	−0.003
q3	0.484	0.008	0.396	0.097	0.519	0.032	0.358
q6	0.637	−0.035	0.231	0.131	0.611	−0.019	0.253
q7	0.888	0.062	0.065	−0.050	0.874	0.044	0.072
q13	0.540	−0.300	0.010	−0.050	0.529	−0.328	0.004
q16	0.496	−0.269	0.057	−0.101	0.518	−0.284	0.026
q18	0.889	0.009	0.005	0.015	0.891	0.000	−0.028
q19	0.760	−0.175	−0.087	−0.020	0.771	−0.149	−0.078
q21	0.715	−0.296	−0.032	0.080	0.692	−0.250	0.015
q39	−0.470	0.101	−0.121	0.268	−0.463	0.158	−0.145
q33	0.207	−0.698	−0.183	0.019	0.176	−0.668	−0.127
q36	−0.045	−0.908	0.003	0.127	−0.018	−0.827	−0.006
q38	0.112	−0.665	0.049	−0.127	0.062	−0.768	0.058
q41	−0.174	0.425	−0.147	0.264	−0.158	0.463	−0.216
q42	0.016	0.823	0.034	0.166	0.063	0.914	0.006
q2	0.201	−0.008	0.598	0.154	0.209	0.060	0.627
q9	−0.041	0.095	0.780	0.062	−0.113	0.050	0.853
q10	0.106	0.070	0.591	−0.230	0.119	−0.048	0.527
q23	−0.023	0.023	0.595	−0.006	−0.021	−0.007	0.603
q27	0.240	−0.219	0.457	−0.107	0.282	−0.231	0.444
q29	0.026	−0.037	0.636	−0.105	0.112	−0.043	0.547
q30	0.198	−0.177	0.420	−0.097	0.142	−0.285	0.438
q11	−0.004	0.158	−0.421	0.395	—	—	—
q12	0.274	0.019	0.463	0.078	—	—	—
q14	0.396	−0.285	0.280	0.026	—	—	—
q15	0.230	−0.306	0.179	−0.324	—	—	—
q17	0.397	−0.437	0.031	0.067	—	—	—

q20	0.027	0.025	0.399	0.393	—	—	—
q26	0.222	−0.346	0.227	0.287	—	—	—
q28	0.211	0.142	0.309	0.325	—	—	—
q34	−0.115	0.407	−0.244	0.179	—	—	—
q35	−0.264	0.513	−0.124	0.221	—	—	—
q37	0.439	0.002	0.398	−0.063	—	—	—
q43	−0.140	−0.419	0.399	0.186	—	—	—
q44	0.024	−0.219	0.344	−0.018	—	—	—

Model 2

Based on initial factor loadings, items were iteratively removed and the structure was reduced from 4 factors to 3. In total, 13 items were removed (see Table 4), which resulted in a 22 item, 3 factor structure with adequate factor loadings and cross-loadings. This model had the following model fit indices: $\chi^2(168) = 404.78$, p = 0; CLI = 0.947; TLI = 0.927; RMSEA = 0.069. In this solution, factor 1 explained 47% of the total variance (eigenvalue = 10.45), factor 2 explained 12% of the total variance (eigenvalue = 2.74), and factor 3 explained 5% of the total variance (eigenvalue = 1.09). This model shows acceptable fit, and was chosen as the final solution for this study.

Scale Inspection

The first factor shows good reliability, with alpha, split half, and greatest lower bound of reliability greater than .93. The second factor has acceptable split half and internal reliability, but it's greatest lower bound is very low. Finally, the third factor has split half reliability, internal consistency, and greatest lower bound between .7 and .8, which indicates acceptable, if low reliability. Based on these statistics, the first factor provides the best measurement of the construct of interest, while factors 2 and 3 present some concern.

```
scale1 <- hr %>%
  select(q1, q3, q6, q7, q13, q16, q18, q19, q21,q39)
scale2 <- hr %>%
```

```r
  select(q33, q36, q38, q41, q42)
scale3 <- hr %>%
  select(q2, q9, q10, q23, q27, q29, q30)


splithalf1 <- splitHalf(scale1) ; splithalf2 <- splitHalf(scale2) ; splithalf3 <- splitHalf(
alpha1 <- psych::alpha(scale1, check.keys = TRUE) ; alpha2 <- psych::alpha(scale2, check.key
glb1 <- glb.algebraic(cor(scale1, use = "pairwise.complete.obs")) ; glb2 <- glb.algebraic(co

data.frame(Factor = c("Factor 1", "Factor 1", "Factor 1", "Factor 2", "Factor 2", "Factor 2"
           `Reliability Statistic` = c("Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
           Statistic = c(round(splithalf1$meanr,2),
                         round(alpha1$total$raw_alpha, 2),
                         round(glb1$glb, 2),
                         round(splithalf2$meanr,2),
                         round(alpha2$total$raw_alpha, 2),
                         round(glb2$glb, 2),
                         round(splithalf3$meanr,2),
                         round(alpha3$total$raw_alpha, 2),
                         round(glb3$glb, 2))) %>%
  gt::gt(groupname_col = "Factor") %>%
  gt::tab_header(title = "Table 5: Reliability Statistics For All Scales")
```

Table 5: Reliability Statistics For All Scales

Reliability.Statistic	Statistic
Factor 1	
Average Split Half	0.93
Alpha	0.93
Greatest Lower Bound	0.93
Factor 2	
Average Split Half	0.82
Alpha	0.85
Greatest Lower Bound	0.41
Factor 3	

Average Split Half	0.77
Alpha	0.75
Greatest Lower Bound	0.79

Once reliability for the final factor solution had been inspected, item responses were investigated. For factor 1, distribution of responses across response options showed that the lowest response options (1, "Strongly Disagree" & 2, "Disagree") were underused for many items, except those which loaded negatively (i.e., reverse coded), like the final item in factor 1, in which the highest response option (6, "Strongly Agree") was used only 7% of the time.

```
item <- read.csv("references/docs/items.csv")


items.1 <-
  data.frame(alpha1$response.freq) %>%
  mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
  pivot_longer(c(X1:X6)) %>%
  mutate(Response =
         factor(case_when(
           name == "X1" ~ "Strongly Disagree",
           name == "X2" ~ "Disagree",
           name == "X3" ~ "Somewhat Disagree",
           name == "X4" ~ "Somewhat Agree",
           name == "X5" ~ "Agree",
           name == "X6" ~ "Strongly Agree"
         ),
         levels = c("Strongly Agree","Agree","Somewhat Agree",
                    "Somewhat Disagree","Disagree",
                    "Strongly Disagree"))) %>%
  ggplot() +
  aes(x = question, y = value, fill = Response) +
  geom_bar(stat = "identity", position = "stack") +
    geom_text(aes(label = paste0(round(value*100,0),"%")),
              position = position_stack(vjust = .5),
```

```r
               size = 8, color = "white") +
  ggokabeito::scale_fill_okabe_ito(
      breaks = c("Strongly Disagree","Disagree","Somewhat Disagree",
                 "Somewhat Agree","Agree","Strongly Agree"
      )) +
    coord_flip() +
    ggforce::facet_col(facets = vars(question),
                       scales = "free_y",
                       space = "free",
                       labeller = label_wrap_gen(width = 100)
    ) +
    labs(title = "Factor 1") +
    theme(
      axis.title = element_blank(),
      strip.background = element_blank(),
      legend.position = "top",
      axis.ticks = element_blank(),
      axis.text = element_blank(),
      panel.background = element_blank(),
      strip.text = element_text(size = 15, vjust = -.2),
      panel.spacing = unit(-.5, "lines"),
      legend.title = element_blank(),
      strip.clip = "off",
      plot.title = element_text(hjust = 0.5, size = 30),
      legend.text = element_text(size = 10)
    )

## Scale 2
items.2 <- data.frame(alpha2$response.freq) %>%
    mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
  pivot_longer(c(X1:X6)) %>%
  mutate(Response =
           factor(case_when(
             name == "X1" ~ "Strongly Disagree",
             name == "X2" ~ "Disagree",
             name == "X3" ~ "Somewhat Disagree",
```

```
            name == "X4" ~ "Somewhat Agree",
            name == "X5" ~ "Agree",
            name == "X6" ~ "Strongly Agree"
          ),
          levels = c("Strongly Agree","Agree","Somewhat Agree",
                     "Somewhat Disagree","Disagree",
                     "Strongly Disagree"))) %>%
ggplot() +
aes(x = question, y = value, fill = Response) +
geom_bar(stat = "identity", position = "stack") +
  geom_text(aes(label = paste0(round(value*100,0),"%")),
            position = position_stack(vjust = .5),
            size = 8, color = "white") +
  ggokabeito::scale_fill_okabe_ito(
    breaks = c("Strongly Disagree","Disagree","Somewhat Disagree",
               "Somewhat Agree","Agree","Strongly Agree"
    )) +
  coord_flip() +
  ggforce::facet_col(facets = vars(question),
                     scales = "free_y",
                     space = "free",
                     labeller = label_wrap_gen(width = 100)
  ) +
  labs(title = "Factor 2") +
  theme(
   axis.title = element_blank(),
    strip.background = element_blank(),
    legend.position = "top",
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    panel.background = element_blank(),
    strip.text = element_text(size = 10),
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )
```

```r
## Scale 3
items.3 <- data.frame(alpha3$response.freq) %>%
mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
  pivot_longer(c(X1:X6)) %>%
  mutate(Response =
           factor(case_when(
             name == "X1" ~ "Strongly Disagree",
             name == "X2" ~ "Disagree",
             name == "X3" ~ "Somewhat Disagree",
             name == "X4" ~ "Somewhat Agree",
             name == "X5" ~ "Agree",
             name == "X6" ~ "Strongly Agree"
           ),
           levels = c("Strongly Agree","Agree","Somewhat Agree",
                      "Somewhat Disagree","Disagree",
                      "Strongly Disagree"
                 ))) %>%
  ggplot() +
  aes(x = question, y = value, fill = Response) +
  geom_bar(stat = "identity", position = "stack") +
    geom_text(aes(label = paste0(round(value*100,0),"%")),
              position = position_stack(vjust = .52),
              size = 3, color = "white") +
    ggokabeito::scale_fill_okabe_ito(
      breaks = c("Strongly Disagree","Disagree","Somewhat Disagree",
                 "Somewhat Agree","Agree","Strongly Agree"
      )) +
    coord_flip() +
    ggforce::facet_col(facets = vars(question),
                       scales = "free_y",
                       space = "free",
                       labeller = label_wrap_gen(width = 100)
    ) +
    labs(title = "Factor 3") +
    theme(
      axis.title = element_blank(),
```

```
    strip.background = element_blank(),
    legend.position = "top",
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    panel.background = element_blank(),
    strip.text = element_text(size = 10),
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )

items.1
```

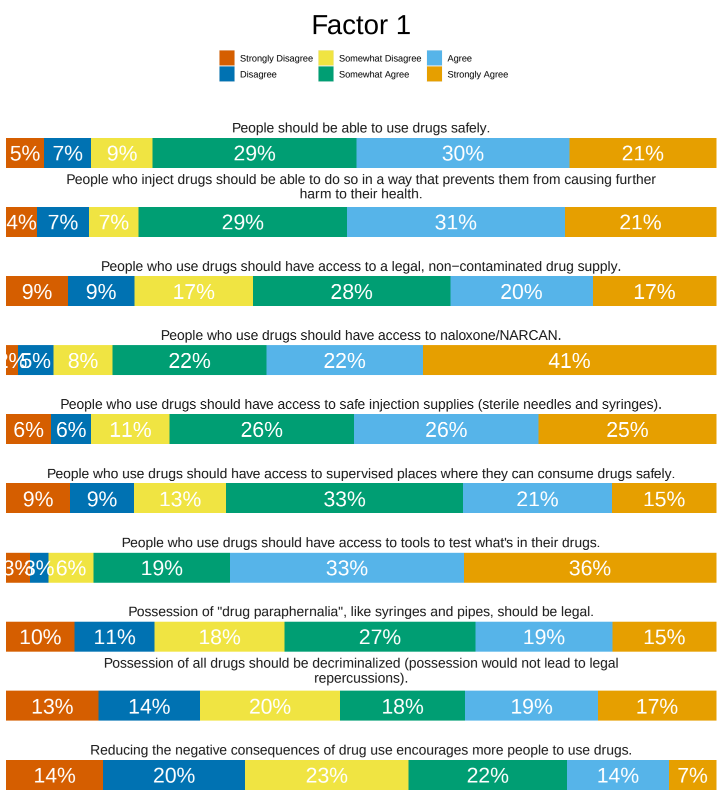


Figure 1: Figure 1: Study 1, Factor 1 Response Distribution

The same was true for both Factors 2 and 3, which showed that extreme response options were underutilized across most items (See Figures 2 & 3).

```
items.2
items.3
```

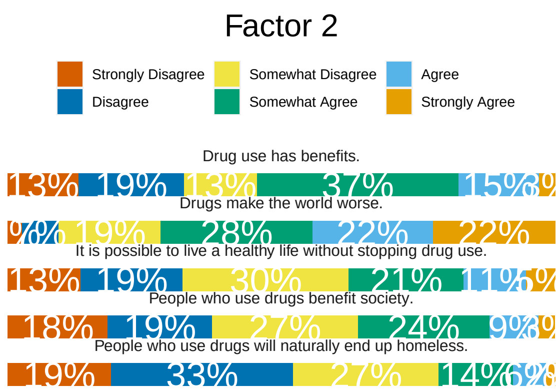


Figure 2: Figure 2: Study 1, Factor 2 Response Distribution

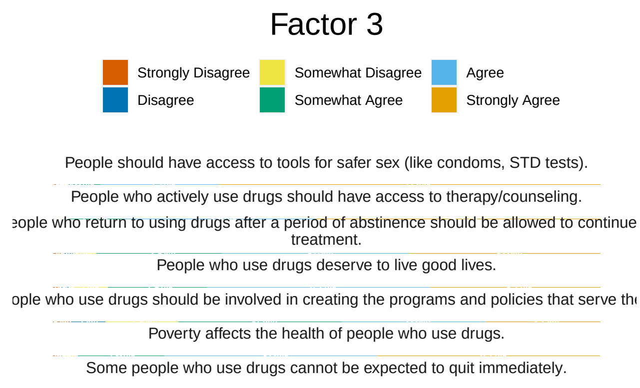


Figure 3: Figure 3: Study 1, Factor 3 Response Distribution

As a result of this response option utilization, the decision was made to reduce from 6 options to 5 for future data collection efforts. The final factor solution was deemed acceptable for confirmatory studies.

Study 2

Study 2 was preregistered, which can be accessed at https://osf.io/t5e2a. In this study, we conducted confirmatory factor analysis using new data based on the factor structure in Study 1. The confirmatory factor analysis was fit using weighted least squares estimation treating categories as ordinal rather than numeric, which was a deviation from the preregistration. We will use Hu & Bentler's (2009) cut offs of .95 for TLI and CFI, and .06 for RMSEA.

Participants

Through Prolific, we recruited 402 participants for Study 2. These participants were 49% men (n = 196), 71% White (n = 282), and most had completed a bachelor's degree (37%, n = 150), and 90% (n = 362) reported no history of SUD diagnosis. Table 6 provides detail about the participants.

```r
df <- read.csv("data/clean/20231020_hr-scale-confirmatory-data.csv")

demos <- df %>%
  select(age, gid1, race, ed, sud_hx)

demos %>%
  tbl_summary(
    missing = "no",
    statistic = list(age = "{mean} ({sd})"),
    digits = list(age ~ c(2,2)),
    sort = list(everything() ~ "frequency"),
    label = list(
      age ~ "Age", gid1 ~ "Gender Identity",
      race ~ "Race/Ethnicity", ed ~ "Highest Level of Education",
      sud_hx ~ "History of Substance Use Disorder Diagnosis")
                ) %>%
  as_gt() %>%
  tab_header(title = "Table 6: Study 2 Participants")
```

Table 6: Study 2 Participants

Characteristic	**N = 402**[1]
Age	36.45 (12.35)
Gender Identity	
Woman	196 (49%)
Man	191 (48%)
Non-Binary, Agender, or Other	13 (3.3%)
Race/Ethnicity	
White	282 (71%)
Asian or Asian American	41 (10%)
Black or African American	33 (8.3%)
Latine or Hispanic	18 (4.5%)

Multiracial	17 (4.3%)
Other or Prefer not to say	8 (2.0%)
Highest Level of Education	
Bachelor's Degree	150 (37%)
High School	122 (30%)
Master's Degree	50 (12%)
Associate's Degree	44 (11%)
Other option not represented here	11 (2.7%)
Professional Degree (J.D., M.D., etc.)	11 (2.7%)
Doctorate Degree (PhD, PsyD, etc.)	5 (1.2%)
Less than High School	4 (1.0%)
Prefer not to say	4 (1.0%)
History of Substance Use Disorder Diagnosis	
No	362 (90%)
Yes	27 (6.7%)
Prefer not to say	8 (2.0%)
I don't know	5 (1.2%)

[1]Mean (SD); n (%)

Confirmator Factory Analysis

The model from Study 1 was fit using confirmatory factor analysis using a weighted least squares estimator and ordinal indicators. This model showed moderate fit based on the selected fit statistics: $\chi^2(206) = 1033.11$, $p > 0.001$; CFI = .97; TLI = .97; SRMR = .27; RMSEA = .10. Based on CFI and TLI, the model had good fit. Based on SRMR and RMSEA, the model shows poor fit. Figure 4 provides factor loadings and covariances for the model.

```
library(lavaan)
library(semPlot)
library(semptools)

hr <- df %>% # N = 301
  select(starts_with("q"), -Q_RecaptchaScore)

# Initial Model Fit
```

```r
model <-  'F1 =~ q1 + q3 + q6 + q7 + q13 + q16 + q18 + q19 + q21 +q39
           F2 =~ q33 + q36 + q38 + q41 + q42
           F3 =~ q2 + q9 + q10 + q23 + q27 + q29 + q30

           F1 ~~ F2
           F1 ~~ F3'

fit <- cfa(
  model,
  data = df,
  ordered = TRUE,
  estimator = "WLS",
  std.lv = TRUE
)


fitmeasures(fit,
            c("cfi", "tli", "srmr", "rmsea", "chisq", "df", "pvalue")) %>%
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_header(
    title = "Table 7: CFA Fit Statistics"
  ) %>%
  fmt_number(columns = 2,
             decimals = 2)
```

Table 7: CFA Fit Statistics

Item	value
cfi	0.97
tli	0.97
srmr	0.27
rmsea	0.10
chisq	1,033.11
df	206.00
pvalue	0.00

```r
plot(set_cfa_layout(
  semPaths(fit, whatLabels="est",
```

```
        sizeMan = 3,
        node.width = 1,
        thresholds = FALSE,
        edge.label.cex = .75,
        style = "ram",
        mar = c(5, 1, 5, 1),
        intercepts = FALSE,
        DoNotPlot = TRUE),
  fcov_curve = 1.75,
  loading_position = .9))
```

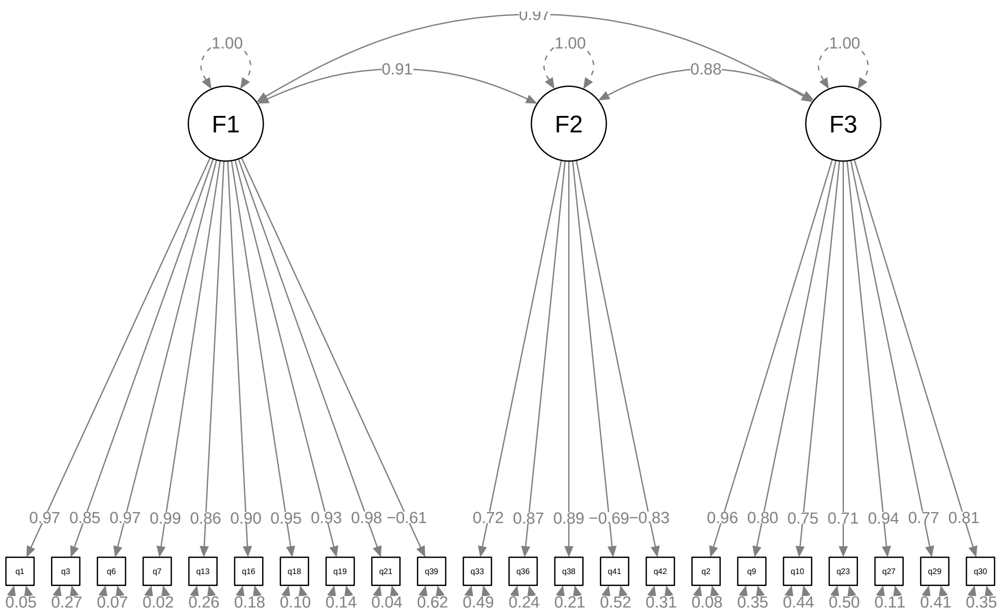


Figure 4: Figure 4: Study 2 CFA

Scale Reliability

Reliability is an issue with this model, as the first factor shows continued adequate reliability, but factors 2 and 3 continue to have poorer reliability. Additionally, reliability would be improved by dropping items, which would shorten factor 2, and make it only 3 items, if not fewer.

```r
scale1 <- hr %>%
  select(q1, q3, q6, q7, q13, q16, q18, q19, q21,q39)
scale2 <- hr %>%
  select(q33, q36, q38, q41, q42)
scale3 <- hr %>%
  select(q2, q9, q10, q23, q27, q29, q30)


splithalf1 <- splitHalf(scale1) ; splithalf2 <- splitHalf(scale2) ; splithalf3 <- splitHalf(
alpha1 <- psych::alpha(scale1, check.keys = TRUE) ; alpha2 <- psych::alpha(scale2, check.key
glb1 <- glb.algebraic(cor(scale1, use = "pairwise.complete.obs")) ; glb2 <- glb.algebraic(co

data.frame(Factor = c("Factor 1", "Factor 1", "Factor 1", "Factor 2", "Factor 2", "Factor 2"
           `Reliability Statistic` = c("Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
           Statistic = c(round(splithalf1$meanr,2),
                         round(alpha1$total$raw_alpha, 2),
                         round(glb1$glb, 2),
                         round(splithalf2$meanr,2),
                         round(alpha2$total$raw_alpha, 2),
                         round(glb2$glb, 2),
                         round(splithalf3$meanr,2),
                         round(alpha3$total$raw_alpha, 2),
                         round(glb3$glb, 2))) %>%
  gt::gt(groupname_col = "Factor") %>%
  gt::tab_header(title = "Table 5: Reliability Statistics For Both Scales")
```

Table 5: Reliability Statistics For Both Scales

Reliability.Statistic	Statistic
Factor 1	
Average Split Half	0.92
Alpha	0.92
Greatest Lower Bound	0.93
Factor 2	
Average Split Half	0.79
Alpha	0.81

Greatest Lower Bound	0.38
Factor 3	
Average Split Half	0.78
Alpha	0.76
Greatest Lower Bound	0.81

Exploratory Factor Analysis

```
hr <- df %>% # N = 301
  select(starts_with("q"), -Q_RecaptchaScore)

hr.ld <-
  hr %>%
  select(-c(q4:q5,
            q31,
            q40,
            q25,
            q45))

hrMatrix <- POLYCHORIC_R(hr.ld,
                         verbose = FALSE)


bart <- cortest.bartlett(hrMatrix, n = 301)

kmo <- KMO(hrMatrix)
```

Based on poor model fit, this same data set was used to conduct a second exploratory factor analysis using a maximum likelihood estimator and an oblimin rotation. This exploratory study was not preregistered. The items which showed initial local dependence were reinspected. In this study, items 8, 24, 22, and 32 showed acceptable levels of local dependence, and were retained. After removing items which showed excessive local dependence, bartlett's sphericity and Kaiser-Meyer-Olkin were checked again, with Kaiser Meyer-Olkin = 0.9291942, and Barlett's sphericity $\chi^2(741) = 8493.3950304$, p = 0.

```r
# Initial Factor Analysis
efa <- EFA.dimensions::EFA(hrMatrix,
                    extraction = "ml",
                    Nfactors = 4,
                    Ncases = 301,
                    rotation = "oblimin",
                    verbose = FALSE)
```

```r
initial <- efa$pattern %>%
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_style(
    locations = list(
      # Factor 1
      cells_body(
      columns = `Factor 1`,
      rows = abs(`Factor 1`) < .30 ),
      # Factor 2
      cells_body(
        columns = `Factor 2`,
        rows = abs(`Factor 2`) < .30
      ),
      # Factor 3
       cells_body(
        columns = `Factor 3`,
        rows = abs(`Factor 3`) < .30
      ),
      # Factor 4
       cells_body(
        columns = `Factor 4`,
        rows = abs(`Factor 4`) < .30
      )),
    style = list(cell_text(color = 'gray'))) %>%
  fmt_number(
    columns = starts_with("Factor"),
    decimals = 3)
```

Initial EFA Model

An initial factor analysis revealed that a four factor solution provided poor fit for the data, with factor 1 explaining 41% of the total variance (eigenvalue = 14.57), factor 2 explaining 8% of the total variance (eigenvalue = 2.8), factor 3 explaining 3% of the total variance (eigenvalue = 1.05), and factor 4 explaining 2.5% of the total variance (eigenvalue = .90). This model had the following model fit indices: $\chi^2(591) = 1457.37$, p = $7.9090106 \times 10^{-75}$; CLI = 0.888; TLI = 0.86; RMSEA = 0.07. In addition to poor fit, there were numerous cross-loadings which exceeded the limit of difference of .2. Table 6 provides full factor loadings.

After removing items with no primary loading or close cross loadings, a three factor solution provided the best fit. At this point, factor 3 consisted of 6 items which explained only 22% of the total variance and had 4 factor loadings less that .55. Between these quantitative reasons and inspecting the qualitative content of the items, they were removed. These were items 2, 9, 28, 29, and 32. The content of these items focused primarily on treatment aspects of harm reduction, where were less directly relevant to constructs of interest.

```
# New Matrix
hrMatrix <-
  POLYCHORIC_R(
    hr %>%
      select(
        # Factor 1
        q1, q6, q7, q16, q18, q19, q21, q13, q39,
        # Factor 2
        q42, q36, q33, q35, q34, q38, q41),
    verbose = FALSE)

efa.final <- EFA.dimensions::EFA(hrMatrix,
                    extraction = "ml",
                    Nfactors = 2,
                    Ncases = 301,
                    rotation = "oblimin",
```

```
                      verbose = FALSE)

final <- efa.final$pattern %>%
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_style(
    locations = list(
      # Factor 1
      cells_body(
      columns = `Factor 1`,
      rows = abs(`Factor 1`) < .30),
      # Factor 2
      cells_body(
        columns = `Factor 2`,
        rows = abs(`Factor 2`) < .30
      )),
    style = list(cell_text(color = 'gray'))) %>%
  fmt_number(
    columns = starts_with("Factor"),
    decimals = 3)
```

Final EFA Model

The resulting model was a two factor model which consisted of 15 items. Factor 1 explained 59% of the variance, and Factor 2 explained 41% of the variance. This model had the following model fit indices: $\chi^2(89) = 241.98$, p $= 4.3949155 \times 10^{-16}$; CLI $= 0.955$; TLI $= 0.94$; RMSEA $= 0.076$, which indicated acceptable fit. Due to the low loading in this study, item "q39" is a candidate for removal following the next round of data collection.

```
left_join(
    initial$`_data`,
    final$`_data`,
    by = c(
      "Item" = "Item"
    )) %>%
```

```
arrange(
  factor(Item,
         levels =
           c(
             # Factor 1
             "q1", "q6", "q7", "q16", "q18", "q19", "q21", "q13", "q39",
             # Factor 2
             "q42", "q36", "q33", "q35", "q34", "q38", "q41"))) %>%
janitor::clean_names() %>%
gt() %>%
tab_style(
  locations = list(
    # Factor 1
    cells_body(
    columns = factor_1_x,
    rows = abs(factor_1_x) < .30),
    # Factor 2
    cells_body(
      columns = factor_2_x,
      rows = abs(factor_2_x) < .30
    ),
    # Factor 3
     cells_body(
      columns = factor_3,
      rows = abs(factor_3) < .30
    ),
    # Factor 4
     cells_body(
      columns = factor_4,
      rows = abs(factor_4) < .30
    ),
     # Factor 1 - Final
    cells_body(
    columns = factor_1_y,
    rows = abs(factor_1_y) < .30),
    # Factor 2 - Final
    cells_body(
      columns = factor_2_y,
      rows = abs(factor_2_y) < .30
```

```r
    )),
    style = list(cell_text(color = 'gray'))) %>%
  fmt_number(
    columns = starts_with("Factor"),
    decimals = 3)  %>%
  fmt_missing() %>%
  tab_header(
    title = "Table 8: Initial and Final Factor Structure for 2 factors"
  ) %>%
  tab_spanner(
    label = "Initial 4 Factor Solution",
    columns = 2:5
  ) %>%
  tab_spanner(
    label = "Final 2 Factor Solution",
    columns = 6:7
  ) %>%
  cols_label(
    factor_1_x  = "Factor 1",
    factor_2_x = "Factor 2",
    factor_3 = "Factor 3",
    factor_4 = "Factor 4",
    factor_1_y = "Factor 1",
    factor_2_y = "Factor 2"
  )
```

Table 8: Initial and Final Factor Structure for 2 factors

	Initial 4 Factor Solution				Final 2 Factor Solution	
item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2
q1	0.868	0.064	−0.015	0.014	0.903	−0.001
q6	0.773	0.182	0.046	0.068	0.852	0.088
q7	0.897	0.094	0.093	−0.064	1.022	0.137
q16	0.674	−0.170	−0.100	−0.137	0.577	−0.202
q18	0.812	−0.056	−0.080	−0.041	0.712	−0.182
q19	0.767	−0.053	−0.124	−0.040	0.719	−0.171
q21	0.745	0.015	−0.290	0.065	0.634	−0.348
q13	0.572	0.042	−0.231	−0.046	0.610	−0.201
q39	−0.395	0.051	−0.032	0.330	−0.542	−0.011

q42	−0.186	0.074	0.710	−0.016	0.043	0.852
q36	0.246	−0.020	−0.655	0.186	−0.050	−0.789
q33	0.281	−0.203	−0.534	−0.061	0.053	−0.674
q35	−0.132	−0.056	0.595	0.221	−0.215	0.605
q34	0.031	−0.097	0.540	0.262	−0.112	0.500
q38	0.450	−0.165	−0.390	−0.187	0.327	−0.513
q41	0.015	−0.217	0.530	0.225	−0.187	0.439
q2	0.148	0.761	−0.079	−0.022	—	—
q3	0.339	0.356	0.074	−0.259	—	—
q8	0.806	0.020	−0.088	−0.026	—	—
q9	0.099	0.670	0.087	−0.176	—	—
q10	0.136	0.333	0.012	−0.385	—	—
q11	−0.272	−0.004	0.020	0.349	—	—
q12	0.204	0.294	−0.042	−0.200	—	—
q14	0.523	0.114	−0.025	−0.247	—	—
q15	0.418	−0.102	−0.198	−0.336	—	—
q17	0.443	0.088	−0.475	0.098	—	—
q20	0.102	0.519	0.259	0.217	—	—
q22	0.305	0.129	−0.046	−0.493	—	—
q23	0.312	0.354	0.220	−0.102	—	—
q24	0.581	0.165	0.184	−0.130	—	—
q26	0.426	0.267	−0.308	0.104	—	—
q27	0.376	0.237	−0.090	−0.348	—	—
q28	0.190	0.458	0.092	0.063	—	—
q29	0.094	0.406	−0.190	−0.206	—	—
q30	0.234	0.219	−0.098	−0.315	—	—
q32	−0.006	0.428	−0.011	−0.388	—	—
q37	0.543	0.221	0.129	−0.105	—	—
q43	−0.177	0.553	−0.484	−0.008	—	—
q44	0.261	0.048	−0.051	−0.196	—	—

The resulting scale consists of the following items:

```
final.items <- hr %>%
      select(
        # Factor 1
        q1, q6, q7, q16, q18, q19, q21, q13, q39,
        # Factor 2
        q42, q36, q33, q35, q34, q38, q41)
```

```r
items <- readr::read_csv("references/codebooks/20231025_hr-scale-codebook.csv")

items %>%
  filter(
    grepl("q", var_label)
    ) %>%
  filter(var_label %in%
           c(# Factor 1
             "q1", "q6", "q7", "q16", "q18", "q19", "q21", "q13", "q39",
             # Factor 2
             "q42", "q36", "q33", "q35", "q34", "q38", "q41")) %>%
  select(-1) %>%
  arrange(
      factor(
    var_label,
    levels = c(# Factor 1
             "q1", "q6", "q7", "q16", "q18", "q19", "q21", "q13", "q39",
             # Factor 2
             "q42", "q36", "q33", "q35", "q34", "q38", "q41"))) %>%
  mutate(
    item =
      case_when(
        var_label %in% c("q42", "q35", "q34", "q41", "q39") ~ paste0(item, "*"),
        TRUE ~ item
      ),
    Factor = c(rep("Factor 1", 9), rep("Factor 2", 7))
  ) %>%
  gt() %>%
  tab_header(
    title = "Table 9: Final Items"
  ) %>%
  cols_label(
    var_label = "Item Label",
    item = "Item Wording"
  ) %>%
  tab_footnote("* indicates reverse coded items")
```

Table 9: Final Items

Item Label	Item Wording
q1	People who inject drugs should be able to do so in a way that prevents them from causing further
q6	People who use drugs should have access to tools to test what's in their drugs.
q7	People who use drugs should have access to safe injection supplies (sterile needles and syringes).
q16	Possession of all drugs should be decriminalized (possession would not lead to legal repercussions).
q18	People who use drugs should have access to supervised places where they can consume drugs safel
q19	People who use drugs should have access to a legal, non-contaminated drug supply.
q21	People should be able to use drugs safely.
q13	Possession of "drug paraphernalia", like syringes and pipes, should be legal.
q39	Reducing the negative consequences of drug use encourages more people to use drugs. *
q42	Drugs make the world worse. *
q36	Drug use has benefits.
q33	It is possible to live a healthy life without stopping drug use.
q35	Using drugs is immoral. *
q34	People who use drugs should be forced into treatment. *
q38	People who use drugs benefit society.
q41	People who use drugs will naturally end up homeless. *

* indicates reverse coded items

Scale Inspection

The first factor shows good reliability, with alpha, split half, and greatest lower bound of reliability greater than .93. The second factor has acceptable split half and internal reliability, but it's greatest lower bound is very low, which continues to be a concern. Overall, between the number of items, EFA fit statistics, and reliability, this model provides better fit and output compared to the initial model.

```
scale1 <- df %>%
  select(q1, q6, q7, q16, q18, q19, q21, q13, q39)
scale2 <- df %>%
  select(q42, q36, q33, q35, q34, q38, q41)


splithalf1 <- splitHalf(scale1) ; splithalf2 <- splitHalf(scale2)
alpha1 <- psych::alpha(scale1, check.keys = TRUE) ; alpha2 <- psych::alpha(scale2, check.key
```

```
glb1 <- glb.algebraic(cor(scale1, use = "pairwise.complete.obs")) ; glb2 <- glb.algebraic(co

data.frame(Factor = c("Factor 1", "Factor 1", "Factor 1", "Factor 2", "Factor 2", "Factor 2"
           `Reliability Statistic` = c("Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
           Statistic = c(round(splithalf1$meanr,2),
                         round(alpha1$total$raw_alpha, 2),
                         round(glb1$glb, 2),
                         round(splithalf2$meanr,2),
                         round(alpha2$total$raw_alpha, 2),
                         round(glb2$glb, 2))) %>%
  gt::gt(groupname_col = "Factor") %>%
  gt::tab_header(title = "Table 10: Reliability Statistics For All Scales")
```

Table 10: Reliability Statistics For All Scales

Reliability.Statistic	Statistic
Factor 1	
Average Split Half	0.91
Alpha	0.92
Greatest Lower Bound	0.92
Factor 2	
Average Split Half	0.86
Alpha	0.85
Greatest Lower Bound	0.46

Item Response Distribution

Both factors show a similar distribution to in the initial study. Some more extreme response options show under use, but across all items response options are used.

```
data.frame(alpha1$response.freq) %>%
  mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
```

```
pivot_longer(c(X1:X5)) %>%
mutate(Response =
         factor(case_when(
           name == "X1" ~ "Strongly disagree",
           name == "X2" ~ "Somewhat disagree",
           name == "X3" ~ "Neither agree nor disagree",
           name == "X4" ~ "Somewhat agree",
           name == "X5" ~ "Strongly Agree"
         ),
           levels = c("Strongly disagree",
                      "Somewhat disagree",
                      "Neither agree nor disagree",
                      "Somewhat agree",
                      "Strongly Agree"))) %>%
ggplot() +
aes(x = question, y = value, fill = Response) +
geom_bar(stat = "identity", position = "stack") +
geom_text(aes(label = paste0(round(value*100,0),"%")),
          position = position_stack(vjust = .5),
          size = 3) +
  coord_flip() +
  ggforce::facet_col(facets = vars(question),
                     scales = "free_y",
                     space = "free",
                     labeller = label_wrap_gen(width = 100)
  ) +
  labs(title = "Factor 1") +
  theme(
   axis.title = element_blank(),
    strip.background = element_blank(),
    legend.position = "top",
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    panel.background = element_blank(),
    strip.text = element_text(size = 10),
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )
```

```
data.frame(alpha2$response.freq) %>%
  mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
  pivot_longer(c(X1:X5)) %>%
  mutate(Response =
  factor(case_when(
    name == "X1" ~ "Strongly disagree",
  name == "X2" ~ "Somewhat disagree",
  name == "X3" ~ "Neither agree nor disagree",
  name == "X4" ~ "Somewhat agree",
  name == "X5" ~ "Strongly Agree"),
  levels = c("Strongly disagree",
             "Somewhat disagree",
             "Neither agree nor disagree",
             "Somewhat agree",
             "Strongly Agree"))) %>%
  ggplot() +
  aes(x = question, y = value, fill = Response) +
  geom_bar(stat = "identity", position = "stack") +
  geom_text(aes(label = paste0(round(value*100,0),"%")),
            position = position_stack(vjust = .5),
            size = 3) +
    coord_flip() +
    ggforce::facet_col(facets = vars(question),
                       scales = "free_y",
                       space = "free",
                       labeller = label_wrap_gen(width = 100)
    ) +
    labs(title = "Factor 2") +
    theme(
     axis.title = element_blank(),
      strip.background = element_blank(),
      legend.position = "top",
      axis.ticks = element_blank(),
      axis.text = element_blank(),
      panel.background = element_blank(),
      strip.text = element_text(size = 10),
```

```
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )
```

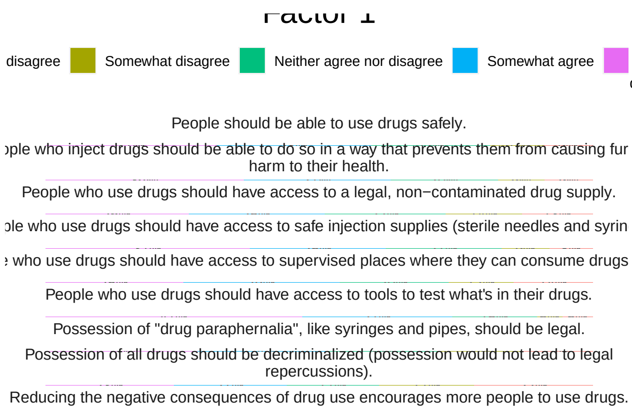


Figure 5: Figure 5: Study 2 EFA, Factor 1 Response Distribution

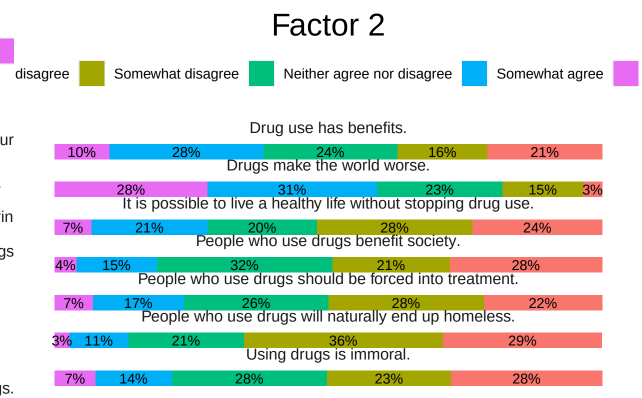


Figure 6: Figure 6: Study 2 EFA, Factor 2 Response Distribution

Study 3

Study 3 was preregistered, which can be accessed at https://osf.io/q4uzv. In this study, we conducted confirmatory factor analysis using new data based on the factor structure in Study 2. The confirmatory factor analysis was fit using weighted least squares estimation treating categories as ordinal rather than numeric, which was a deviation from the preregistration. We will use Hu & Bentler's (2009) cut offs of .95 for TLI and CFI, and .06 for RMSEA.

Participants

Through Prolific, we recruited 402 participants for Study 2. These participants were 52% men (n = 205), 70% White (n = 277), and most had completed a bachelor's degree (41%, n =

161), and 88% reported no history of SUD diagnosis. Table 6 provides detail about the participants.

```
df <- read.csv("data/clean/20231025_hr-scale-confirmatory-data2.csv")


demos <- df %>%
  select(age, gid1, race, ed, sud_hx)

demos %>%
  tbl_summary(
    missing = "no",
    statistic = list(age = "{mean} ({sd})"),
    digits = list(age ~ c(2,2)),
    sort = list(everything() ~ "frequency"),
    label = list(
      age ~ "Age", gid1 ~ "Gender Identity",
      race ~ "Race/Ethnicity", ed ~ "Highest Level of Education",
      sud_hx ~ "History of Substance Use Disorder Diagnosis")
              ) %>%
  as_gt() %>%
  tab_header(title = "Table 11: Study 3 Participants")
```

Table 11: Study 3 Participants

Characteristic	**N = 393**[1]
Age	37.03 (11.59)
Gender Identity	
Man	205 (52%)
Woman	174 (44%)
Non-Binary, Agender, or Other	14 (3.6%)
Race/Ethnicity	
White	277 (70%)
Asian or Asian American	41 (10%)
Black or African American	35 (8.9%)
Latine or Hispanic	28 (7.1%)
Multiracial	12 (3.1%)
Highest Level of Education	
Bachelor's Degree	161 (41%)
High School	129 (33%)

Associate's Degree	46 (12%)
Master's Degree	34 (8.7%)
Professional Degree (J.D., M.D., etc.)	9 (2.3%)
Other option not represented here	5 (1.3%)
Doctorate Degree (PhD, PsyD, etc.)	4 (1.0%)
Less than High School	3 (0.8%)
History of Substance Use Disorder Diagnosis	
No	346 (88%)
Yes	40 (10%)
I don't know	5 (1.3%)
Prefer not to say	2 (0.5%)

[1]Mean (SD); n (%)

Confirmatory Factor Analysis

The model developed in study 2 was fit to study 3's data using confirmatory factor analysis with least square estimation and ordingal indicators. The model showed acceptablve fit on most indicators: χ^2 (103) = 314.68, p < .001; CFI = .98, TLI = .97; RMSEA = .07; SRMR = .19. Figure 7 provides factor loadings and covariances for the model.

```r
# Import data
df <- read.csv("data/clean/20231025_hr-scale-confirmatory-data2.csv")

# Select items
hr <- df %>% # N = 301
  select(starts_with("q"), -Q_RecaptchaScore)

# Study 2 Model Fit
model <-  'F1 =~ q21 + q1 + q6 + q7 + q16 + q18 + q19 + q13 + q39
           F2 =~ q36 + q42 + q33 + q35 + q34 + q38 + q41

           F1 ~~ F2'

fit <- cfa(
  model,
  data = df,
```

```r
  ordered = TRUE,
  estimator = "WLS",
  std.lv = TRUE
)


fitmeasures(fit,
            c("cfi", "tli", "srmr", "rmsea", "chisq", "df", "pvalue")) %>%
  as_tibble(rownames = "Item") %>%
  gt() %>%
  tab_header(
    title = "Table 12: Study 3 CFA Fit Statistics"
  ) %>%
  fmt_number(columns = 2,
             decimals = 2)
```

Table 12: Study 3 CFA Fit Statistics

Item	value
cfi	0.98
tli	0.97
srmr	0.19
rmsea	0.07
chisq	314.68
df	103.00
pvalue	0.00

```r
plot(set_cfa_layout(
  semPaths(fit, whatLabels="est",
        sizeMan = 3,
        node.width = 1,
        thresholds = FALSE,
        edge.label.cex = .75,
        style = "ram",
        mar = c(5, 1, 5, 1),
        intercepts = FALSE,
        DoNotPlot = TRUE),
  fcov_curve = 1.75,
  loading_position = .9))
```

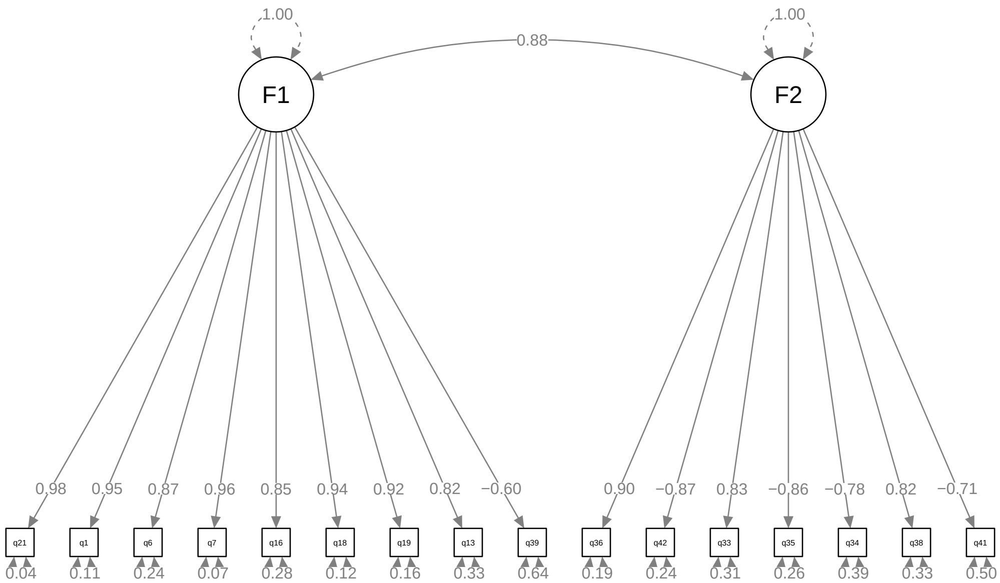


Figure 7: Figure 7: Study 3 CFA

Scale Inspection

The first factor shows good reliability, with alpha, split half, and greatest lower bound of reliability greater than .93. The second factor has acceptable split half and internal reliability, but it's greatest lower bound is very low, which continues to be a concern. Overall, between the number of items, EFA fit statistics, and reliability, this model provides better fit and output compared to the initial model.

```
scale1 <- df %>%
  select(q1, q6, q7, q16, q18, q19, q21, q13, q39)
scale2 <- df %>%
  select(q42, q36, q33, q35, q34, q38, q41)


splithalf1 <- splitHalf(scale1) ; splithalf2 <- splitHalf(scale2)
alpha1 <- psych::alpha(scale1, check.keys = TRUE) ; alpha2 <- psych::alpha(scale2, check.key
```

```r
glb1 <- glb.algebraic(cor(scale1, use = "pairwise.complete.obs")) ; glb2 <- glb.algebraic(co

data.frame(Factor = c("Factor 1", "Factor 1", "Factor 1", "Factor 2", "Factor 2", "Factor 2"
           `Reliability Statistic` = c("Average Split Half", "Alpha", "Greatest Lower Bound"
                                       "Average Split Half", "Alpha", "Greatest Lower Bound"
           Statistic = c(round(splithalf1$meanr,2),
                         round(alpha1$total$raw_alpha, 2),
                         round(glb1$glb, 2),
                         round(splithalf2$meanr,2),
                         round(alpha2$total$raw_alpha, 2),
                         round(glb2$glb, 2))) %>%
  gt::gt(groupname_col = "Factor") %>%
  gt::tab_header(title = "Table 10: Reliability Statistics For All Scales")
```

Table 10: Reliability Statistics For All Scales

Reliability.Statistic	Statistic
Factor 1	
Average Split Half	0.91
Alpha	0.92
Greatest Lower Bound	0.92
Factor 2	
Average Split Half	0.86
Alpha	0.86
Greatest Lower Bound	0.46

Item Response Distribution

Both factors show a similar distribution to in the initial study. Some more extreme response options show under use, but across all items response options are used.

```r
data.frame(alpha1$response.freq) %>%
  mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
```

```
pivot_longer(c(X1:X5)) %>%
mutate(Response =
         factor(case_when(
           name == "X1" ~ "Strongly disagree",
           name == "X2" ~ "Somewhat disagree",
           name == "X3" ~ "Neither agree nor disagree",
           name == "X4" ~ "Somewhat agree",
           name == "X5" ~ "Strongly Agree"
         ),
           levels = c("Strongly disagree",
                      "Somewhat disagree",
                      "Neither agree nor disagree",
                      "Somewhat agree",
                      "Strongly Agree"))) %>%
ggplot() +
aes(x = question, y = value, fill = Response) +
geom_bar(stat = "identity", position = "stack") +
geom_text(aes(label = paste0(round(value*100,0),"%")),
          position = position_stack(vjust = .5),
          size = 3) +
  coord_flip() +
  ggforce::facet_col(facets = vars(question),
                     scales = "free_y",
                     space = "free",
                     labeller = label_wrap_gen(width = 100)
  ) +
  labs(title = "Factor 1") +
  theme(
   axis.title = element_blank(),
    strip.background = element_blank(),
    legend.position = "top",
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    panel.background = element_blank(),
    strip.text = element_text(size = 10),
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )
```

```
data.frame(alpha2$response.freq) %>%
  mutate(item = rownames(.)) %>%
  left_join(item %>%
              select(qname, question),
            by = c("item" = "qname")) %>%
  pivot_longer(c(X1:X5)) %>%
  mutate(Response =
  factor(case_when(
    name == "X1" ~ "Strongly disagree",
  name == "X2" ~ "Somewhat disagree",
  name == "X3" ~ "Neither agree nor disagree",
  name == "X4" ~ "Somewhat agree",
  name == "X5" ~ "Strongly Agree"),
  levels = c("Strongly disagree",
             "Somewhat disagree",
             "Neither agree nor disagree",
             "Somewhat agree",
             "Strongly Agree"))) %>%
  ggplot() +
  aes(x = question, y = value, fill = Response) +
  geom_bar(stat = "identity", position = "stack") +
  geom_text(aes(label = paste0(round(value*100,0),"%")),
            position = position_stack(vjust = .5),
            size = 3) +
    coord_flip() +
    ggforce::facet_col(facets = vars(question),
                       scales = "free_y",
                       space = "free",
                       labeller = label_wrap_gen(width = 100)
    ) +
    labs(title = "Factor 2") +
    theme(
     axis.title = element_blank(),
      strip.background = element_blank(),
      legend.position = "top",
      axis.ticks = element_blank(),
      axis.text = element_blank(),
      panel.background = element_blank(),
      strip.text = element_text(size = 10),
```

```
    panel.spacing = unit(-.5, "lines"),
    legend.title = element_blank(),
    strip.clip = "off",
    plot.title = element_text(hjust = 0.5, size = 20)
  )
```

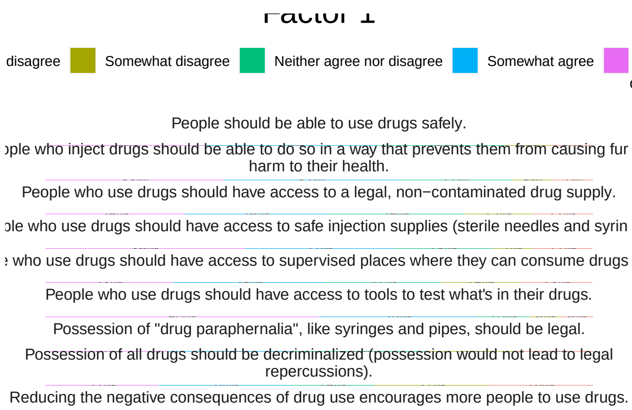


Figure 8: Figure 8: Study 3, Strategies Subscale Response Distribution

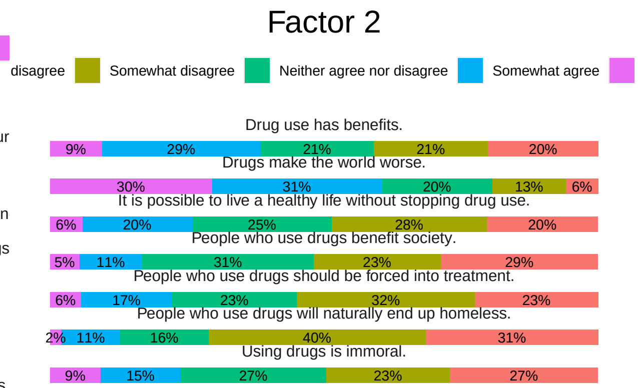


Figure 9: Figure 9: Study 3, Principles Subscale Response Distribution