

# ReCentering Psych Stats: Psychometrics

Lynette H. Bikos, PhD, ABPP (she/her)

02 Apr 2024



# Contents

<b>1</b>	<b>Introduction</b>	<b>21</b>
1.1	What to expect in each chapter . . . . .	21
1.2	Strategies for Accessing and Using this OER . . . . .	22
1.3	If You are New to R . . . . .	22
1.3.1	Base R . . . . .	22
1.3.2	R Studio . . . . .	23
1.3.3	R Hygiene . . . . .	23
1.3.3.1	Everything is documented in the .rmd file . . . . .	24
1.3.3.2	File organization . . . . .	24
1.3.3.3	Chunks . . . . .	24
1.3.3.4	Packages . . . . .	25
1.3.3.5	Knitting . . . . .	25
1.3.4	tRoubleshooting in R maRkdown . . . . .	25
1.3.5	stRategies for success . . . . .	26
1.3.6	Resources for getting staRted . . . . .	26
<b>2</b>	<b>Questionnaire Construction: The Fundamentals</b>	<b>29</b>
2.1	Navigating this Lesson . . . . .	29
2.1.1	Learning Objectives . . . . .	29
2.1.2	Planning for Practice . . . . .	30
2.1.3	Readings & Resources . . . . .	30
2.2	Components of the Questionnaire . . . . .	30
2.3	What Improves (or Threatens) Response Rates and Bias? . . . . .	32
2.3.1	Should Likert-type scales include a midpoint? . . . . .	32
2.3.2	Should <i>continuous rating scales</i> be used in surveys? . . . . .	33
2.3.3	Should Likert-type response options use an ascending or descending order? .	34

2.3.4	Should surveys include negatively worded items? . . . . .	35
2.4	Construct-specific guidance . . . . .	36
2.5	Surveying in the Online Environment . . . . .	37
2.6	In my Surveys . . . . .	38
2.6.1	Demographics and Background Information . . . . .	38
2.6.2	Survey Order . . . . .	41
2.6.3	Forced Responses . . . . .	41
2.7	Practice Problems . . . . .	41
<b>3</b>	<b>Be a QualTRIXter</b>	<b>43</b>
3.1	Navigating this Lesson . . . . .	43
3.1.1	Learning Objectives . . . . .	43
3.1.2	Planning for Practice . . . . .	43
3.1.3	Readings & Resources . . . . .	44
3.1.4	Packages . . . . .	44
3.2	Research Vignette . . . . .	44
3.3	Qualtrics Essentials . . . . .	45
3.4	Qual-TRIX . . . . .	46
3.5	Even moRe, particularly relevant to iRb . . . . .	47
3.6	intRavenous Qualtrics . . . . .	48
3.6.1	The Codebook . . . . .	50
3.6.2	Using data from an exported Qualtrics .csv file . . . . .	51
3.6.3	Tweaking Data Format . . . . .	51
3.7	Practice Problems . . . . .	51
<b>4</b>	<b>Psychometric Validity: Basic Concepts</b>	<b>53</b>
4.1	Navigating this Lesson . . . . .	53
4.1.1	Learning Objectives . . . . .	53
4.1.2	Planning for Practice . . . . .	54
4.1.3	Readings & Resources . . . . .	54
4.1.4	Packages . . . . .	54
4.2	Research Vignette . . . . .	54
4.3	Fundamentals of Validity . . . . .	59
4.4	Validity Criteria . . . . .	60

<b>CONTENTS</b>	<b>5</b>
4.4.1 Content Validity . . . . .	60
4.4.2 Face Validity: The “Un”validity . . . . .	61
4.4.3 Criterion-Related Validity . . . . .	61
4.4.4 Construct Validity . . . . .	63
4.4.5 Internal Consistency . . . . .	63
4.4.6 Structural Validity . . . . .	63
4.4.6.1 Exploratory Factor Analysis . . . . .	63
4.4.6.2 Confirmatory Factor Analysis . . . . .	64
4.4.7 Experimental Interventions . . . . .	64
4.4.8 Convergent and Discriminant Validity . . . . .	64
4.4.8.1 Determining Statistically Significant Differences Between Correlations	68
4.4.8.2 Multitrait-Multimethod Matrix . . . . .	71
4.4.9 Incremental Validity . . . . .	71
4.4.10 Considering the Individual and Social Consequences of Testing . . . . .	75
4.5 Factors Affecting Validity Coefficients . . . . .	75
4.6 Practice Problems . . . . .	77
4.6.1 Problem #1: Play around with this simulation. . . . .	78
4.6.2 Problem #2: Conduct the reliability analysis selecting different variables. . . . .	78
4.6.3 Problem #3: Try something entirely new. . . . .	78
4.6.4 Grading Rubric . . . . .	78
4.7 Homeworked Example . . . . .	79
4.7.1 Check and, if needed, format data . . . . .	79
4.7.2 Create a correlation matrix that includes the instrument-of-interest and the variables that will have varying degrees of relation . . . . .	81
4.7.3 With convergent and discriminant validity in mind, interpret the validity coefficients; this should include an assessment about whether the correlation coefficients (at least two different pairings) are statistically significantly different from each other. . . . .	82
4.7.4 With at least three variables, evaluate the degree to which the instrument demonstrates incremental validity (this should involve two regression equations and their statistical comparison) . . . . .	86
<b>5 Reliability</b>	<b>89</b>
5.1 Navigating this Lesson . . . . .	89
5.1.1 Learning Objectives . . . . .	89
5.1.2 Planning for Practice . . . . .	89

5.1.3	Readings & Resources . . . . .	90
5.1.4	Packages . . . . .	90
5.2	Defining Reliability . . . . .	91
5.2.1	Begins with Classical Test Theory (CTT) . . . . .	91
5.2.2	Why are we concerned with reliability? Error! . . . . .	91
5.2.3	The Reliability Coefficient . . . . .	92
5.3	Research Vignette . . . . .	93
5.4	A Parade of Reliability Coefficients . . . . .	97
5.4.1	Reliability Options for a Single Administration . . . . .	98
5.4.1.1	Split-half reliability . . . . .	98
5.4.1.2	Alpha coefficients . . . . .	102
5.4.1.3	Omega . . . . .	106
5.4.1.4	Some summary statements about reliability from single administrations . . . . .	112
5.4.2	Reliability Options for Two or more Administrations . . . . .	112
5.4.2.1	Test-retest of total scores . . . . .	112
5.4.2.2	Test retest recap . . . . .	115
5.4.3	Interrater Reliability . . . . .	116
5.4.3.1	Cohen's kappa . . . . .	116
5.4.3.2	Intraclass correlation (ICC) . . . . .	118
5.5	What do we do with these coefficients? . . . . .	119
5.5.1	Corrections for attenuation . . . . .	119
5.5.2	Predicting true scores (and their CIs) . . . . .	119
5.5.3	How do I keep it all straight? . . . . .	120
5.6	Practice Problems . . . . .	120
5.6.1	Problem #1: Play around with this simulation. . . . .	120
5.6.2	Problem #2: Use the data from the live ReCentering Psych Stats survey. . . . .	121
5.6.3	Problem #3: Try something entirely new. . . . .	122
5.6.4	Grading Rubric . . . . .	122
5.7	Homeworked Example . . . . .	122
5.7.1	Check and, if needed, format the data . . . . .	123
5.7.2	Calculate and report the alpha coefficient for a total scale score and subscales (if the scale has them) . . . . .	124
5.7.3	Subscale alphas . . . . .	126

5.7.4 Calculate and report $t$ and $h$ . . . . .	129
5.7.5 With these two determine what proportion of the variance is due to all the factors, error, and $g$ . . . . .	133
5.7.6 Calculate total and subscale scores. . . . .	133
5.7.7 Describe other reliability estimates that would be appropriate for the measure you are evaluating. . . . .	133
<b>6 Item Analysis for Educational Achievement Tests (Exams)</b>	<b>135</b>
6.1 Navigating this Lesson . . . . .	135
6.1.1 Learning Objectives . . . . .	135
6.1.2 Planning for Practice . . . . .	136
6.1.3 Readings & Resources . . . . .	136
6.1.4 Packages . . . . .	136
6.2 Research Vignette . . . . .	136
6.3 Item Analysis in the Educational/Achievement Context . . . . .	136
6.3.1 And now a quiz! Please take it. . . . .	137
6.4 Item Difficulty . . . . .	139
6.4.1 Percent passing . . . . .	139
6.4.2 Several factors prevent .50 from being the ideal difficulty level . . . . .	139
6.5 Item Discrimination . . . . .	140
6.5.1 Index of Discrimination . . . . .	141
6.5.2 Application of Item Difficulty and Discrimination . . . . .	141
6.6 In the <i>psych</i> Package . . . . .	146
6.6.1 A Mini-Introduction to IRT . . . . .	151
6.7 Closing Thoughts on Developing Measures in the Education/Achievement Context .	159
6.8 Practice Problems . . . . .	160
<b>7 Item Analysis for Likert Type Scale Construction</b>	<b>161</b>
7.1 Navigating this Lesson . . . . .	161
7.1.1 Learning Objectives . . . . .	161
7.1.2 Planning for Practice . . . . .	162
7.1.3 Readings & Resources . . . . .	162
7.1.4 Packages . . . . .	162
7.2 Introducing Item Analysis for Survey Development . . . . .	163
7.2.1 Workflow for Item Analysis . . . . .	163

7.3	Research Vignette . . . . .	164
7.4	Step I: Corrected item-total correlations . . . . .	168
7.4.1	Data Prep . . . . .	169
7.4.2	Calculating Item-Total Correlation Coefficients . . . . .	171
7.5	Step II: Correlating Items with Other Scale Totals . . . . .	175
7.6	Step III: Interpreting and Writing up the Results . . . . .	177
7.7	A Conversation with Dr. Szymanski . . . . .	178
7.8	Practice Problems . . . . .	178
7.8.1	Problem #1: Play around with this simulation. . . . .	179
7.8.2	Problem #2: Use raw data from the ReCentering Psych Stats survey on Qualtrics. . . . .	179
7.8.3	Problem #3: Try something entirely new. . . . .	180
7.8.4	Grading Rubric . . . . .	180
7.9	Bonus Reel: . . . . .	181
7.10	Homeworked Example . . . . .	182
7.10.1	Check and, if needed, format and score data . . . . .	183
7.10.2	Report alpha coefficients and average inter-item correlations for the total and subscales . . . . .	184
7.10.3	Produce and interpret corrected item-total correlations for total and subscales, separately . . . . .	184
7.10.3.1	All Course Evaluation Items . . . . .	184
7.10.3.2	Valued-by-Student Subscale . . . . .	186
7.10.3.3	Traditional Pedagogy Items . . . . .	187
7.10.3.4	Socially Responsive Pedagogy Items . . . . .	189
7.10.4	Produce and interpret correlations between the individual items of a given subscale and the subscale scores of all other subscales . . . . .	190
7.10.4.1	Valued-by-the-Student Items . . . . .	190
7.10.5	Traditional Pedagogy Items . . . . .	193
7.10.5.1	Socially Responsive Pedagogy Items . . . . .	196
7.10.6	APA style results section with table . . . . .	198
7.10.7	Explanation to grader . . . . .	198

<b>8 Principal Components Analysis</b>	<b>201</b>
8.1 Navigating this Lesson . . . . .	201
8.1.1 Learning Objectives . . . . .	201
8.1.2 Planning for Practice . . . . .	202
8.1.3 Readings & Resources . . . . .	202
8.1.4 Packages . . . . .	203
8.2 Exploratory Principal Components Analysis . . . . .	203
8.2.1 Some Framing Ideas (in very lay terms) . . . . .	203
8.3 PCA Workflow . . . . .	204
8.4 Research Vignette . . . . .	206
8.5 Working the Vignette . . . . .	209
8.5.1 Three Diagnostic Tests to Evaluate the Appropriateness of the Data for Component-or-Factor Analysis . . . . .	211
8.5.1.1 Is my sample adequate for PCA? . . . . .	211
8.5.1.2 Are there correlations among the variables that are large enough to be analyzed? . . . . .	212
8.5.1.3 Is there multicollinearity or singularity in my data? . . . . .	213
8.5.2 Principal Components Analysis . . . . .	214
8.5.3 Specifying the Number of Components . . . . .	219
8.5.3.1 Quick recap of how to evaluate the # of components we extracted .	228
8.5.4 Component Rotation . . . . .	228
8.5.4.1 Orthogonal rotation . . . . .	229
8.5.4.2 Oblique rotation . . . . .	234
8.5.5 Component Scores . . . . .	240
8.6 APA Style Results . . . . .	242
8.7 Back to the FutuRe: The relationship between PCA and item analysis . . . . .	243
8.7.1 Calculating and Extracting Item-Total Correlation Coefficients . . . . .	244
8.7.1.1 Corrected item-total correlations from the <i>psych::alpha()</i> . . . . .	244
8.7.1.2 Correlating items with other subscale totals . . . . .	250
8.7.1.3 Extracting values, binding them together, and joining the files . .	250
8.7.1.4 Interpreting the result . . . . .	254
8.8 Practice Problems . . . . .	255
8.8.1 Problem #1: Play around with this simulation. . . . .	255
8.8.2 Problem #2: Conduct a PCA with another simulated set of data in the OER.	255

8.8.3	Problem #3: Try something entirely new. . . . .	255
8.8.4	Grading Rubric . . . . .	256
8.9	Homeworked Example . . . . .	256
8.9.1	Check and, if needed, format data . . . . .	257
8.9.2	Conduct and interpret the three diagnostic tests to determine if PCA is appropriate as an analysis (KMO, Bartlett's, determinant) . . . . .	257
8.9.2.1	KMO . . . . .	257
8.9.2.2	Bartlett's . . . . .	258
8.9.2.3	Determinant . . . . .	259
8.9.3	Determine how many components to extract (e.g., scree plot, eigenvalues, theory) . . . . .	259
8.9.4	Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions . . . . .	262
8.9.5	Conduct an oblique extraction and rotation with a minimum of two different factor extractions . . . . .	267
8.9.6	Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest . . . . .	273
8.9.7	APA style results section with table and figure of one of the solutions . . . . .	273
8.9.8	Explanation to grader . . . . .	276
<b>9</b>	<b>Principal Axis Factoring</b>	<b>277</b>
9.1	Navigating this Lesson . . . . .	277
9.1.1	Learning Objectives . . . . .	277
9.1.2	Planning for Practice . . . . .	278
9.1.3	Readings & Resources . . . . .	278
9.1.4	Packages . . . . .	278
9.2	Exploratory Factor Analysis (with a quick contrast to PCA) . . . . .	279
9.3	PAF Workflow . . . . .	280
9.4	Research Vignette . . . . .	282
9.5	Working the Vignette . . . . .	284
9.5.1	Data Prep . . . . .	285
9.5.1.1	Three Diagnostic Tests to Evaluate the Appropriateness of the Data for Component (or Factor)Analysis . . . . .	286
9.5.1.2	Is my sample adequate for PAF? . . . . .	286
9.5.1.3	Are there correlations among the variables that are big enough to be analyzed? . . . . .	287

9.5.1.4	Is there multicollinearity or singularity in my data? . . . . .	288
9.5.1.5	APA Style Summary So Far . . . . .	288
9.5.2	Principal Axis Factoring (PAF) . . . . .	289
9.5.2.1	Specifying the number of factors . . . . .	294
9.5.2.2	Quick recap of how to evaluate the # of factors we extracted . . . . .	303
9.5.3	Factor Rotation . . . . .	304
9.5.3.1	Orthogonal rotation . . . . .	304
9.5.3.2	Oblique rotation . . . . .	309
9.5.4	Factor Scores . . . . .	314
9.6	APA Style Results . . . . .	317
9.6.1	Comparing FA and PCA . . . . .	318
9.7	Going Back to the Future: What, then, is Omega? . . . . .	319
9.8	Comparing PFA to Item Analysis and PCA . . . . .	325
9.9	Practice Problems . . . . .	327
9.9.1	Problem #1: Play around with this simulation. . . . .	328
9.9.2	Problem #2: Conduct a PCA with the Szymanski and Bissonette [2020] research vignette that was used in prior lessons. . . . .	328
9.9.3	Problem #3: Try something entirely new. . . . .	328
9.9.4	Grading Rubric . . . . .	328
9.10	Homeworked Example . . . . .	329
9.10.1	Check and, if needed, format data . . . . .	329
9.10.2	Conduct and interpret the three diagnostic tests to determine if PCA is appropriate as an analysis (KMO, Bartlett's, determinant) . . . . .	330
9.10.2.1	KMO . . . . .	330
9.10.2.2	Bartlett's . . . . .	331
9.10.2.3	Determinant . . . . .	332
9.10.3	Determine how many components to extract (e.g., scree plot, eigenvalues, theory) . . . . .	332
9.10.4	Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions . . . . .	335
9.10.5	Conduct an oblique extraction and rotation with a minimum of two different factor extractions . . . . .	341
9.10.6	Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest . . . . .	347
9.10.7	APA style results section with table and figure of one of the solutions . . . . .	348
9.10.8	Explanation to grader . . . . .	350

<b>10 CFA: First Order Models</b>	<b>353</b>
10.1 Navigating this Lesson . . . . .	353
10.1.1 Learning Objectives . . . . .	353
10.1.2 Planning for Practice . . . . .	354
10.1.3 Readings & Resources . . . . .	354
10.1.4 Packages . . . . .	355
10.2 Two Broad Categories of Factor Analysis: Exploratory and Confirmatory . . . . .	355
10.2.1 Common to Both Exploratory and Confirmatory Approaches . . . . .	355
10.2.2 Differences between EFA and CFA . . . . .	356
10.2.3 On the relationship between EFA and CFA . . . . .	357
10.3 Exploring a Standard CFA Model . . . . .	357
10.3.1 Model Identification for CFA . . . . .	359
10.3.2 Selecting Indicators/Items for a Reflective Measurement . . . . .	360
10.4 CFA Workflow . . . . .	360
10.4.1 CFA in <i>lavaan</i> Requires Fluency with the Syntax . . . . .	362
10.4.2 Differing Factor Structures . . . . .	363
10.5 Research Vignette . . . . .	364
10.5.1 Modeling the GRMSAAW as Unidimensional . . . . .	371
10.5.1.1 Interpreting the Output . . . . .	375
10.5.1.2 Model Test <i>User</i> Model: . . . . .	376
10.5.1.3 Model Test <i>Baseline</i> Model . . . . .	377
10.5.1.4 Incremental Fit Indices (User versus Baseline Models) . . . . .	377
10.5.1.5 Loglikelihood and Information Criteria . . . . .	378
10.5.1.6 Root Mean Square Error of Approximation . . . . .	378
10.5.1.7 Standardized Root Mean Square Residual . . . . .	378
10.5.2 Modeling the GRMSAAW as a First-Order, 4-factor model . . . . .	388
10.5.2.1 Specifying and Running the Model . . . . .	388
10.5.2.2 Interpretation . . . . .	393
10.6 Model Comparison . . . . .	396
10.7 A concluding thought . . . . .	399
10.8 Practice Problems . . . . .	400
10.8.1 Problem #1: Play around with this simulation. . . . .	400
10.8.2 Problem #2: Use simulated data from other lessons. . . . .	401
10.8.3 Problem #3: Try something entirely new. . . . .	401

<i>CONTENTS</i>	13
-----------------	----

<b>11 Additional Simulations</b>	<b>403</b>
11.1 iBelong Scale . . . . .	403
11.2 Identity Threat . . . . .	410
11.3 Anti-Racism Behavioral Inventory . . . . .	415



# BOOK COVER

This open education resource is available in two formats:

- Formatted as an [html book](#) via GitHub Pages
- As a [PDF](#)

All materials used in creating this OER are available at its [GitHub repo](#).

As a perpetually-in-progress, open education resource, feedback is always welcome. This IRB-approved (SPU IRB #202102010R, no expiration) [Qualtrics-hosted survey](#) includes formal rating scales, open-ended text boxes, and a portal for uploading attachments (e.g., marked up PDFs). You are welcome to complete only the portions that are relevant to you.

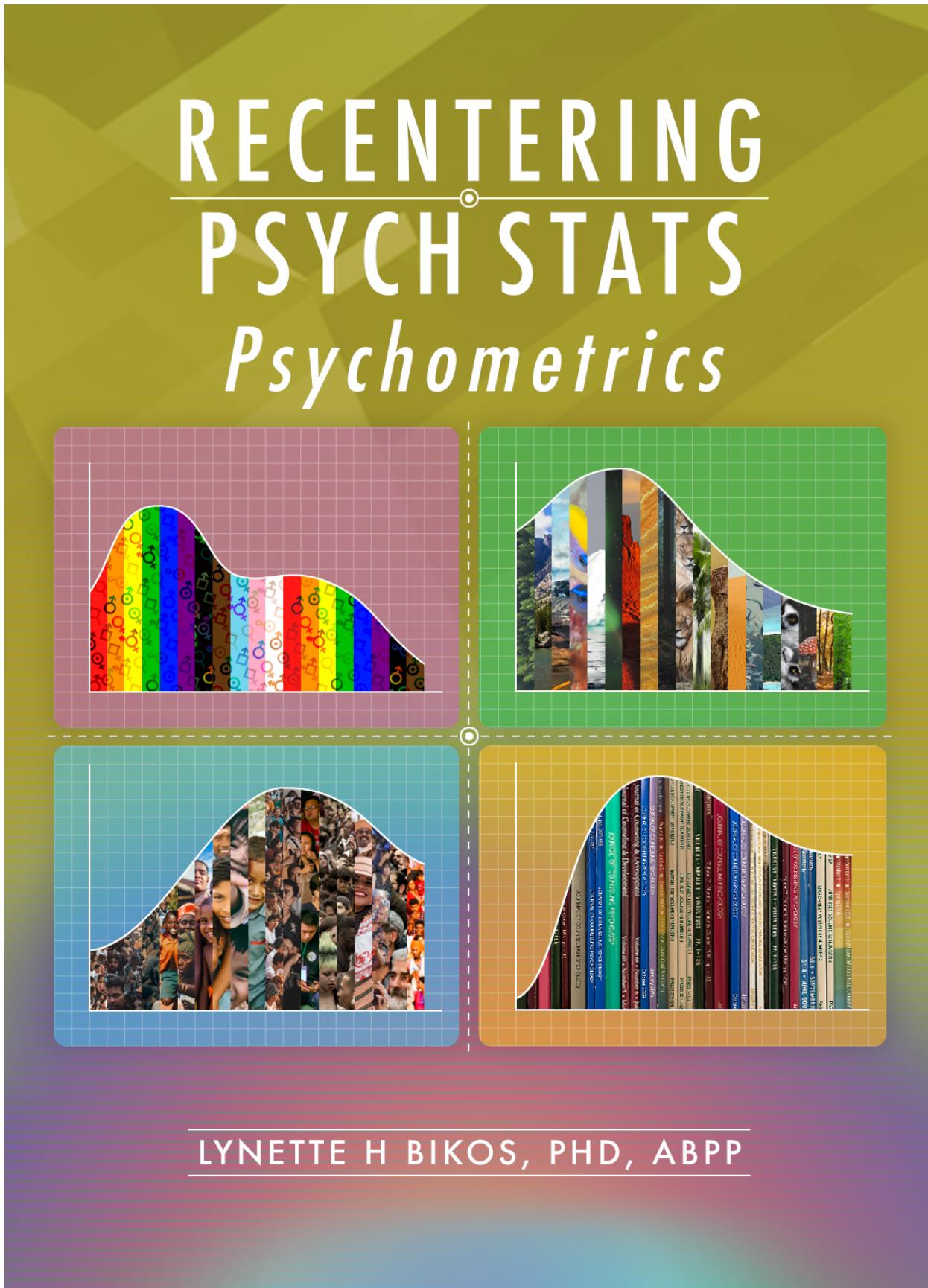


Figure 1: An image of the book cover. It includes four quadrants of non-normal distributions representing gender, race/ethnicity, sustainability/global concerns, and journal articles

# PREFACE

If you are viewing this document, you should know that this is a book-in-progress. Early drafts are released for the purpose teaching my classes and gaining formative feedback from a host of stakeholders. The document was last updated on 02 Apr 2024. Emerging volumes on other statistics are posted on the [ReCentering Psych Stats](#) page at my research team's website.

## [Screencasted Lecture Link](#)

To *center* a variable in regression means to set its value at zero and interpret all other values in relation to this reference point. Regarding race and gender, researchers often center male and White at zero. Further, it is typical that research vignettes in statistics textbooks are similarly seated in a White, Western (frequently U.S.), heteronormative, framework. The purpose of this project is to create a set of open educational resources (OER) appropriate for doctoral and post-doctoral training that contribute to a socially responsive pedagogy – that is, it contributes to justice, equity, diversity, and inclusion.

Statistics training in doctoral programs are frequently taught with fee-for-use programs (e.g., SPSS/AMOS, SAS, MPlus) that may not be readily available to the post-doctoral professional. In recent years, there has been an increase and improvement in R packages (e.g., *psych*, *lavaan*) used for in analyses common to psychological research. Correspondingly, many graduate programs are transitioning to statistics training in R (free and open source). This is a challenge for post-doctoral psychologists who were trained with other software. This OER will offer statistics training with R and be freely available (specifically in a GitHub repository and posted through GitHub Pages) under a Creative Commons Attribution - Non Commercial - Share Alike license [CC BY-NC-SA 4.0].

Training models for doctoral programs in health service psychology are commonly scholar-practitioner, scientist-practitioner, or clinical-scientist. An emerging model, the *scientist-practitioner-advocacy* training model, incorporates social justice advocacy so that graduates are equipped to recognize and address the sociocultural context of oppression and unjust distribution of resources and opportunities [[Mallinckrodt et al., 2014](#)]. In statistics textbooks, the use of research vignettes engages the learner around a tangible scenario for identifying independent variables, dependent variables, covariates, and potential mechanisms of change. Many students recall examples in Field's [[2012](#)] popular statistics text: Viagra to teach one-way ANOVA, beer goggles for two-way ANOVA, and bushtucker for repeated measures. What if the research vignettes were more socially responsive?

In this OER, research vignettes will be from recently published articles where:

- the author's identity is from a group where scholarship is historically marginalized (e.g.,

- BIPOC, LGBTQ+, LMIC[low-middle income countries]),
- the research is responsive to issues of justice, equity, inclusion, diversity,
- the lesson's statistic is used in the article, and
- there is sufficient information in the article to simulate the data for the chapter example(s) and practice problem(s); or it is publicly available.

In training for multicultural competence, the saying, “A fish doesn’t know that it’s wet” is often used to convey the notion that we are often unaware of our own cultural characteristics. In recent months and years, there has been an increased awakening to the institutional and systemic racism that our systems are perpetuating. Queuing from the water metaphor, I am hopeful that a text that is recentered in the ways I have described can contribute to *changing the water* in higher education and in the profession of psychology.

## Copyright with Open Access

This book is published under a a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license: CC BY-SA 4.0.

A [GitHub open-source repository](#) contains all of the text and source code for the book, including data and images.

# ACKNOWLEDGEMENTS

As a doctoral student at the University of Kansas (1992-1996), I learned that “a foreign language” was a graduation requirement. *Please note that as one who studies the intersections of global, vocational, and sustainable psychology, I regret that I do not have language skills beyond English.* This could have been met with credit from high school, but my rural, mid-Missouri high school did not offer such classes. This requirement would have typically been met with courses taken during an undergraduate program – but my non-teaching degree in the University of Missouri’s School of Education was exempt from this. The requirement could have also been met with a computer language (FORTRAN, C++) – but I did not have any of those either. There was a tiny footnote on my doctoral degree plan that indicated that a 2-credit course, “SPSS for Windows” would substitute for the language requirement. Given that it was taught by my one of my favorite professors, I readily signed up. As it turns out, Samuel B. Green, PhD, was using the course to draft chapters in the textbook [[Green and Salkind, 2017](#)] that has been so helpful for so many. Unfortunately, Drs. Green (1947 - 2018) and Salkind (1947 - 2017) are no longer with us. I have worn out numerous versions of their text. Another favorite text of mine has been Dr. Barbara Byrne’s [[2016](#)], “Structural Equation Modeling with AMOS.” I loved the way she worked through each problem and paired it with a published journal article, so that the user could see how the statistical evaluation fit within the larger project/article. I took my tea-stained text with me to a workshop she taught at APA and was proud of the signature she added to it. Dr. Byrne created SEM texts for a number of statistical programs (e.g., LISREL, EQS, MPlus). As I was learning R, I wrote Dr. Byrne, asking if she had an edition teaching SEM/CFA with R. She promptly wrote back, saying that she did not have the bandwidth to learn a new statistics package. We lost Dr. Byrne in December 2020. I am so grateful to these role models for their contributions to my statistical training. I am also grateful for the doctoral students who have taken my courses and are continuing to provide input for how to improve the materials.

The inspiration for training materials that re\*center statistics and research methods came from the [Academics for Black Survival and Wellness Initiative](#). This project, co-founded by Della V. Mosley, Ph.D., and Pearis L. Bellamy, M.S., made clear the necessity and urgency for change in higher education and the profession of psychology.

At very practical levels, I am indebted to SPU’s Library, and more specifically, SPU’s Education, Technology, and Media Department. Assistant Dean for Instructional Design and Emerging Technologies, R. John Robertson, MSc, MCS, has offered unlimited consultation, support, and connection. Senior Instructional Designer in Graphics & Illustrations, Dominic Wilkinson, designed the logo and bookcover. Psychology and Scholarly Communications Librarian, Kristin Hoffman, MLIS, has provided consultation on topics ranging from OERS to citations. I am also indebted to Associate Vice President, Teaching and Learning at Kwantlen Polytechnic University, Rajiv Jhangiani, PhD. Dr. Jhangiani’s text [[2019](#)] was the first OER I ever used and I was grateful for

his encouraging conversation.

Financial support for this project has been provided the following:

- *Call to Action on Equity, Inclusion, Diversity, Justice, and Social Responsivity Request for Proposals* grant from the Association of Psychology Postdoctoral and Internship Centers (2021-2022).
- *Diversity Seed Grant*, Office of Inclusive Excellence and Advisory Council for Diversity and Reconciliation (ACDR), Seattle Pacific University.
- *ETM Open Textbook & OER Development Funding*, Office of Education, Technology, & Media, Seattle Pacific University.

# Chapter 1

## Introduction

[Screencasted Lecture Link](#)

### 1.1 What to expect in each chapter

This textbook is intended as *applied*, in that a primary goal is to help the scientist-practitioner-advocate use a variety of statistics in research problems and *writing them up* for a program evaluation, dissertation, or journal article. In support of that goal, I try to provide just enough conceptual information so that the researcher can select the appropriate statistic (i.e., distinguishing between when ANOVA is appropriate and when regression is appropriate) and assign variables to their proper role (e.g., covariate, moderator, mediator).

This conceptual approach does include occasional, step-by-step, *hand-calculations* (only we calculate them arithmetically in R) to provide a *visceral feeling* of what is happening within the statistical algorithm that may be invisible to the researcher. Additionally, the conceptual review includes a review of the assumptions about the characteristics of the data and research design that are required for the statistic. Statistics can be daunting, so I have worked hard to establish a *workflow* through each analysis. When possible, I include a flowchart that is referenced frequently in each chapter and assists the researcher keep track of their place in the many steps and choices that accompany even the simplest of analyses.

As with many statistics texts, each chapter includes a *research vignette*. Somewhat unique to this resource is that the vignettes are selected from recently published articles. Each vignette is chosen with the intent to meet as many of the following criteria as possible:

- the statistic that is the focus of the chapter was properly used in the article,
- the author's identity is from a group where scholarship is historically marginalized (e.g., BIPOC, LGBTQ+, LMIC [low middle income countries]),
- the research has a justice, equity, inclusion, diversity, and social responsibility focus and will contribute positively to a social justice pedagogy, and
- the data is available in a repository or there is sufficient information in the article to simulate the data for the chapter example(s) and practice problem(s).

In each chapter we employ *R* packages that will efficiently calculate the statistic and the dashboard of metrics (e.g., effect sizes, confidence intervals) that are typically reported in psychological science.

## 1.2 Strategies for Accessing and Using this OER

There are a number of ways you can access this resource. You may wish to try several strategies and then select which works best for you. I demonstrate these in the screencast that accompanies this chapter.

1. Simply follow along in the .html formatted document that is available on via GitHug Pages, and then
  - open a fresh .rmd file of your own, copying (or retyping) the script and running it
2. Locate the original documents at the [GitHub repository](#). You can
  - open them to simply take note of the “behind the scenes” script
  - copy/download individual documents that are of interest to you
  - fork a copy of the entire project to your own GitHub site and further download it (in its entirety) to your personal workspace. The [GitHub Desktop app](#) makes this easy!
3. Listen to the accompanying lectures (I think sound best when the speed is 1.75). The lectures are being recorded in Panopto and should include the closed captioning.
4. Provide feedback to me! If you fork a copy to your own GitHub repository, you can
  - open up an editing tool and mark up the document with your edits,
  - start a discussion by leaving comments/questions, and then
  - sending them back to me by committing and saving. I get an e-mail notifying me of this action. I can then review (accepting or rejecting) them and, if a discussion is appropriate, reply back to you.

## 1.3 If You are New to R

R can be overwhelming. Jumping right into advanced statistics might not be the easiest way to start. However, in these chapters, I provide complete code for every step of the process, starting with uploading the data. To help explain what R script is doing, I sometimes write it in the chapter text; sometimes leave hashtags-comments in the chunks; and, particularly in the accompanying screencasted lectures, try to take time to narrate what the R script is doing.

I've found that, somewhere on the internet, there's almost always a solution to what I'm trying to do. I am frequently stuck and stumped and have spent hours searching the internet for even the tiniest of things. When you watch my videos, you may notice that in my R studio, there is a “scRiptuRe” file. I takes notes on the solutions and scripts here – using keywords that are meaningful to me so that when I need to repeat the task, I can hopefully search my own prior solutions and find a fix or a hint.

### 1.3.1 Base R

The base program is free and is available here: <https://www.r-project.org/>

Because R is already on my machine (and because the instructions are sufficient), I will not walk through the installation, but I will point out a few things.

- Follow the instructions for your operating system (Mac, Windows, Linux)
- The “cran” (I think “cranium”) is the *Comprehensive R Archive Network*. In order for R to run on your computer, you have to choose a location. Because proximity is somewhat related to processing speed, select one that is geographically “close to you.”
- You will see the results of this download on your desktop (or elsewhere if you chose to not have it appear there) but you won’t ever use R through this platform.

### 1.3.2 R Studio

*R Studio* is the desktop application I work in R. It’s a separate download. Choose the free, desktop, option that is appropriate for your operating system: <https://www.rstudio.com/products/RStudio/>

- Upper right window: Includes several tabs; we frequently monitor the
  - Environment: it lists the *objects* that are available to you (e.g., dataframes)
- Lower right window: has a number of helpful tabs.
  - Files: Displays the file structure in your computer’s environment. Make it a practice to (a) organize your work in small folders and (b) navigating to that small folder that is holding your project when you are working on it.
  - Packages: Lists the packages that have been installed. If you navigate to it, you can see if it is “on.” You can also access information about the package (e.g., available functions, examples of script used with the package) in this menu. This information opens in the Help window.
  - Viewer and Plots are helpful, later, when we can simultaneously look at our output and still work on our script.
- Primary window
  - R Studio runs in the background(in the console). Very occasionally, I can find useful troubleshooting information here.
  - More commonly, I open my R Markdown document so that it takes the whole screen and I work directly, right here.
- *R Markdown* is the way that many analysts write *script*, conduct analyses, and even write up results. These are saved as .rmd files.
  - In R Studio, open an R Markdown document through File/New File/R Markdown
  - Specify the details of your document (title, author, desired output)
  - In a separate step, SAVE this document [File/Save] into a NEW FILE FOLDER that will contain anything else you need for your project (e.g., the data).
  - *Packages* are at the heart of working in R. Installing and activating packages require writing script.

### 1.3.3 R Hygiene

Many initial problems in R can be solved with good R hygiene. Here are some suggestions for basic practices. It can be tempting to “skip this.” However, in the first few weeks of class, these are the solutions I am presenting to my students.

### 1.3.3.1 Everything is documented in the .rmd file

Although others do it differently, everything is in my .rmd file. That is, for uploading data and opening packages I write the code in my .rmd file. Why? Because when I read about what I did hours or years later, I have a permanent record of very critical things like (a) where my data is located, (b) what version I was using, and (c) what package was associated with the functions.

### 1.3.3.2 File organization

File organization is a critical key to this:

- Create a project file folder.
- Put the data file in it.
- Open an R Markdown file.
- Save it in the same file folder.
- When your data and .rmd files are in the same folder (not your desktop, but a shared folder), they can be connected.

### 1.3.3.3 Chunks

The R Markdown document is an incredible tool for integrating text, tables, and analyses. This entire OER is written in R Markdown. A central feature of this is “chunks.”

The easiest way to insert a chunk is to use the INSERT/R command at the top of this editor box. You can also insert a chunk with the keyboard shortcut: CTRL/ALT/i

“Chunks” start and end with those three tic marks and will show up in a shaded box, like this:

```
#hashtags let me write comments to remind myself what I did
#here I am simply demonstrating arithmetic (but I would normally be running code)
2021 - 1966
```

```
## [1] 55
```

Each chunk must open and close. If one or more of your tic marks get deleted, your chunk won’t be read as such and your script will not run. The only thing in the chunks should be script for running R; you can hashtag-out script so it won’t run.

Although unnecessary, you can add a brief title for the chunk in the opening row, after the “r.” These create something of a table of contents of all the chunks – making it easier to find what you did. You can access them in the “Chunks” tab at the bottom left of R Studio. If you wish to knit a document, you cannot have identical chunk titles.

You can put almost anything you want in the space outside of tics. Syntax for simple formatting in the text areas (e.g., using italics, making headings, bold, etc.) is found here: [https://rmarkdown.rstudio.com/authoring\\_basics.html](https://rmarkdown.rstudio.com/authoring_basics.html)

### 1.3.3.4 Packages

As scientist-practitioners (and not coders), we will rely on *packages* to do our work for us. At first you may feel overwhelmed about the large number of packages that are available. Soon, though, you will become accustomed to the ones most applicable to our work (e.g., psych, tidyverse, lavaan, apaTables).

Researchers treat packages differently. In these lectures, I list all the packages we will use in an opening chunk that asks R to check to see if the package is installed, and if not, installs it.

```
if (!require(psych)) {
  install.packages("psych")
}
```

```
## Loading required package: psych
```

To make a package operable, you need to open it through the library. This process must be repeated each time you restart R. I don't open the package (through the "library(package\_name)") command until it is time to use it. Especially for new users, I think it's important to connect the functions with the specific packages.

```
# install.packages ('psych')
library(psych)
```

If you type in your own "install.packages" code, hashtag it out once it's been installed. It is problematic to continue to re-run this code .

### 1.3.3.5 Knitting

An incredible feature of R Markdown is its capacity to *knit* to HTML, powerpoint, or word. If you access the .rmd files for this OER, you can use annotate or revise them to suit your purposes. If you redistribute them, though, please honor the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License with a citation.

## 1.3.4 tRoubleshooting in R maRkdown

Hiccups are normal. Here are some ideas that I have found useful in getting unstuck.

- In an R script, you must have everything in order – Every. Single. Time.
  - All the packages have to be in your library and activated; if you restart R, you need to reload each package.
  - If you open an .rmd file and want a boxplot, you cannot just scroll down to that script. You need to run any *prerequisite* script (like loading the package, importing data, putting the data in the global environment, etc.)
  - Do you feel lost? clear your global environment (broom) and start at the top of the R script. Frequent, fresh starts are good.

- Your .rmd file and your data need to be stored in the same file folder. These should be separate for separate projects, no matter how small.
- Type any warnings you get into a search engine. Odds are, you'll get some decent hints in a manner of seconds. Especially at first, these are common errors:
  - The package isn't loaded (if you restarted R, you need to reload your packages)
  - The .rmd file has been saved yet, or isn't saved in the same folder as the data
  - Errors of punctuation or spelling
- Restart R (it's quick – not like restarting your computer)
- If you receive an error indicating that a function isn't working or recognized, and you have loaded the package, type the name of the package in front of the function with two colons (e.g., psych::describe(df)). If multiple packages are loaded with functions that have the same name, R can get confused.

### 1.3.5 stRategies for success

- Engage with R, but don't let it overwhelm you.
  - The *mechanical is also the conceptual*. Especially when it is *simpler*, do try to retype the script into your own .rmd file and run it. Track down the errors you are making and fix them.
  - If this stresses you out, move to simply copying the code into the .rmd file and running it. If you continue to have errors, you may have violated one of the best practices above (Is the package loaded? Are the data and .rmd files in the same place? Is all the prerequisite script run?).
  - Still overwhelmed? Keep moving forward by downloading a copy of the .rmd file that accompanies any given chapter and just “run it along” with the lecture. Spend your mental power trying to understand what each piece does. Then select a practice problem that is appropriate for your next level of growth.
- Copy script that works elsewhere and replace it with your datafile, variables, etc.
- The learning curve is steep, but not impossible. Gladwell[2008] reminds us that it takes about 10,000 hours to get GREAT at something (2,000 to get reasonably competent). Practice. Practice. Practice.
- Updates to R, R Studio, and the packages are NECESSARY, but can also be problematic. It could very well be that updates cause programs/script to fail (e.g., “X has been deprecated for version X.XX”). Moreover, this very well could have happened between my distribution of these resources and your attempt to use it. My personal practice is to update R, R Studio, and the packages a week or two before each academic term.
- Embrace your downward dog. Also, walk away, then come back.

### 1.3.6 Resources for getting staRted

R for Data Science: <https://r4ds.had.co.nz/>

R Cookbook: <http://shop.oreilly.com/product/9780596809164.do>

R Markdown homepage with tutorials: <https://rmarkdown.rstudio.com/index.html>

R has cheatsheets for everything, here's one for R Markdown: <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

R Markdown Reference guide: <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

Using R Markdown for writing reproducible scientific papers: <https://libscie.github.io/rmarkdown-workshop/handout.html>

LaTeX equation editor: <https://www.codecogs.com/latex/eqneditor.php>



# Chapter 2

## Questionnaire Construction: The Fundamentals

### [Screencasted Lecture Link](#)

The focus of this chapter is on the technical issues of constructing a survey. I found this lesson to be more of a struggle to prepare than I expected. Why? There is a great deal of lore about what increases response rates and participation. Yet, research over the years, has both supported, contradicted, or not addressed these claims. One example is where to include “sensitive items.” Historically, textbook authors have recommended that these should come last so that respondents would be engaged in the process and be more willing to complete the survey [Krathwohl, 2009, Rowley, 2014]. Yet, research has shown that this has not held up in employee groups [Roberson and Sundstrom, 1990] nor among members of the National Association of Social Workers [Robert G. Green et al., 2000].

Given these contradictions, this lecture starts with the overall structure of a survey. The core of the lecture focuses on recent, evidence-based support for item-level decisions. I briefly discuss construct-specific guidance and discuss specific considerations for the on-line environment. I then close by addressing some of the decisions that I routinely make in survey construction and provide my rationale for why. Because this lesson occurs at the beginning of a text on psychometrics – this “skips over and around” reliability and validity. These important issues will be addressed in subsequent clessons.

### 2.1 Navigating this Lesson

There is just under one hour of lecture.

While the majority of R objects and data you will need are created within the R script that sources the chapter, there are a few resources that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER’s [introduction](#)

### 2.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Outline the overall structure/components of a questionnaire,
- Articulate test construction myths (e.g., location of sensitive items, “requirement” to have reverse scored items) and their evidence-based solutions (when they have them)
- List elements to consider when the questionnaire is administered online

### 2.1.2 Planning for Practice

This is a two-part lesson on questionnaire construction. After the second lesson, a detailed suggestion for practice will be provided that lists criteria for creating and piloting a survey of your own.

### 2.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

- Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, 56(10), 15–23. <https://doi.org/10.1002/pfi.21727>
- Chyung, S. Y., Barkin, J. R., & Shamsy, J. A. (2018a). Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys. *Performance Improvement*, 57(3), 16–25. <https://doi.org/10.1002/pfi.21749>
- Chung, S. Y., Kennedy, M., & Campbell, I (2018b). Evidence-based survey design: The use of ascending or descending order of Likert-type response options. *Performance Improvement*, 57(9), 9-16. <https://doi.org/10.1002/pfi.21800>
- Chyung, S. Y., Swanson, I., Roberts, K., & Hankinson, A. (2018c). Evidence-Based Survey Design: The Use of Continuous Rating Scales in Surveys. *Performance Improvement*, 57(5), 38–48. <https://doi.org/10.1002/pfi.21763>
  - Finding the Chyung et al. series was like finding a pot of gold! They provide empirical support for guiding choices about survey construction. And they are current! If you don't have time to read them in detail, I recommend you scan them and archive them for future reference.

## 2.2 Components of the Questionnaire

Let's start by examining the components of a questionnaire and the general guidelines for their construction[[Colton and Covert, 2015](#), [Pershing and Pershing, 2001](#)]:

### Title

- reflect the content of the instrument

- be concisely worded
- be written in language easily understood by the respondents
- should not be offensive or off-putting
- should be formatted clearly at the top/beginning of the document

### Introductory Statement

- include a brief summary of the instrument's purpose
- contain an appropriate statement concerning the confidentiality of the respondent's information (informed consent)
- be motivating such that respondents are inspired/willing to complete the items
- specify the approximate amount of time required to complete the instrument

### Directions

- complete, unambiguous, concise
- written at a language level appropriate to the respondents
- tell the respondents how to return the instrument once they have completed it (surprisingly, in Qualtrics, this is also important; submission requires hitting that last little “→”)

### Items

- discussed throughout this textbook

### Closing Statement

- thank the participants for their participation
- remind participants that their information is valuable and perhaps remind about
  - next steps or follow-up
  - confidentiality

### Overall Structure/Look

- should be coherent with an easy-to-follow layout
- professional appearance
  - not crowded, plenty of white space
  - avoiding a “slick look”
  - numbering and headings to provide a sense of progress
  - breaks between every 4-6 questions (or shading alternate items)
  - in a sense, inviting and “easy on the eye”

Pershing and Pershing [2001] reviewed 50 *reactionnaires* that were used by training evaluators at a “prestigious medical school.” Their purpose was to determine the degree to which the survey design adhered to the recommendations. The results suggested that:

- 72% did not include an introductory statement; an additional 16% were “minimal”
- 78% had no closing statement
- 30% had no directions; another 54% of directions were “minimal”
- 8% were professional in appearance

In summary, the formatting of the reactionnaires were not designed in a way that would maximize respondent engagement. In turn, we might expect this to threaten the psychometric reliability and validity.

## 2.3 What Improves (or Threatens) Response Rates and Bias?

When we design survey instruments based on our own preference rather than research-based evidence, we may get less than optimal data. Chyung et al. [2018b] reviewed the five steps [Schwarz and Oyserman, 2001] that survey respondents engage when answering structured, closed-ended survey items.

1. Interpreting the question.
2. Retrieving information from their memory.
3. Integrating the retrieved information with the item prompt.
4. Selecting one of the given response options.
5. Editing the answer for reasons of social desirability.

Chyung and colleagues appear to be starting such a systematic review. What follows are their evidence based evaluations regarding some of the most common questions about questionnaire construction.

### 2.3.1 Should Likert-type scales include a midpoint?

Likert-type scales, named after Rensis Likert, include a set of questions or statements that can be responded to with a consistent set of response options. The response options are frequently scaled with intensity options ranging from 2 to 11; four and five point “agreement” scales str shown below. There are many variants of response options (i.e., frequency, like me, degree of stressfulness). Regarding the issue of a *midpoint* (“neutral” or “neither disagree nor agree”), Chyung et al. [2017] reviewed the literature. Examining their article, we can see variants of Likert-style scaling for a scale of agreement. They look something like this:

---

Type						
No midpoint (4 pt)	Strongly Disagree	Disagree	<i>skipped</i>	Agree	Strongly Agree	
Midpoint (5 pt)	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree	

---

Chyung and colleagues quickly suggest that the question is not “Should I use a midpoint?” but rather “When should I use a midpoint?”

The article is more detailed, but essentially, a midpoint is appropriate when:

- the measurement scale is interval (instead of ordinal; this is a good statistical property to have)
- the question content is such that the midpoint is a *true* midpoint and not a point for hedging or avoiding

If a true midpoint is impossible, then consider adding an option such as “I don’t know” or “It depends.” If the “I don’t know” option is used, the researcher needs a plan for recoding the data as missing and having a plan for managing missingness

### 2.3.2 Should *continuous rating scales* be used in surveys?

First, let’s consider the distinction between *discrete*, *continuous*, and *numerical* scales. Figure 4 in the Chyung, Swanson, Roberts, and Hankinson [2018a] article illustrate the major differences and some variations.

- **Discrete** scales are Likert-type scales that range between 2 and 11 *discrete* options. Classically, respondents pick *words* (e.g., pain rated as *no pain*, *mild*, *moderate*, *severe*, *extreme*, *worst pain possible*).
  - Six-point discrete rating scales result in a collection of six *ordered values*.
  - The measurement scale for discrete scales is *ordinal*.
  - Ordinal scales should be analyzed with non-parametric statistical procedures, however parametric approaches can be used if the data are normally distributed and there is a mid-point.
- **Continuous** scales allow respondents to indicate a response anywhere within a given range – usually by marking a place on a horizontal line on a continuum of a minimum of 100 points. There are no discrete categories defined by words or numbers.
  - Continuous scales result in precise numbers (e.g., 26 or 26.8 if the scale is 0 to 100).
  - The measurement scale for continuous scales is *interval*.
  - Interval scales can be evaluated with parametric statistics.
  - *Visual analog scales (VAS; aka graphic rating scales, GRS)* are another variant of continuous rating scales if they allow the participants to make “anywhere on the line.” Some VAS scales have verbal descriptors to guide the marking; some have numbers (hence, *numerical response scales*). In Qualtrics there is a slider option that serves this function.

Which is better? The mixed results are summarized in Chyung et al’s [2018a] Table 1. With a focus on the types of research I encounter in my program, here is my take-away:

- Continuous scales provide better data (i.e., more precise/full information, more likely to be normally distributed, better reliability) for statistical analysis.
  - *Caveat:* If the response scale on a Likert scale is increased to 11, there is a better chance to have normally distributed responses.
  - *Caveat:* When “simple descriptives” are desired (e.g., histograms, frequency distributions) the discrete scale may be the best choice.

- Discrete and continuous options (including sliders) are easy to use, except in the case where respondents complete the surveys on mobile devices.
  - *Caveat:* There has been more missing data with sliders (compared to radio buttons).
  - *Caveat:* Respondents are more likely to change their responses on sliders. If this means there is greater accuracy or more careful responding, this is desirable.
- In both circumstances adding “don’t know,” “prefer not to respond,” or “not applicable” may improve the validity of the responses.
  - *Caveat:* When using one of these response options, plan ahead for coding as missing and managing the missingness.

### 2.3.3 Should Likert-type response options use an ascending or descending order?

Let’s first look at the difference between ascending and descending order [Chyung et al., 2018b]:

Type	Ascending			Descending		
<b>Ascending</b>	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree	
<b>Descending</b>	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	

In the consideration of the choice between ascending/descending, we are concerned with *response-order effects*. Let’s first examine these conceptually/theoretically.

**Recency effect** is the tendency of survey respondents to select the options that they see at the end of the response-option list. This is expected when options are presented orally (e.g., during interviews, people tend to choose from the last-offered options).

**Primacy effect** is the survey respondents’ tendency to select the options that are presented at the beginning of the response-option list. This is expected when options are presented visually. For example, people tend to choose among the first-presented categories in self-administered written survey questionnaires.

- *Left-sided selection bias* occurs when respondents read text from left-to-right and are more inclined to select options from the left.
- *Satisficing theory* occurs when individuals seek solutions that are “simply satisfactory” so as to minimize psychological costs. Thus, respondents may
  - select the first option that seems “reasonable enough”,
  - select the “I don’t know” response, or
  - randomly select one of the options.
- *Acquiescence bias* is the tendency for respondents to agree with the statement provided—aka yea-saying bias (e.g., being polite).

- Closely related is *social-desirability bias*, the tendency for respondents to select among the options they think are more socially acceptable or desirable (instead of true responses).
- In surveys, this generally is selecting *agree* or *strongly agree*.

Considering these response biases together, Chyung et al. suggest that when the response options are presented in descending order (*Strongly agree, Agree, Neutral, Disagree, Strongly disagree*), respondents would (theoretically) see a positive option immediately on the left side of the response scale and perceive it to be socially desirable and satisfactory. As a result, they may to select it without having to spend more time to choose a more accurate response. After reviewing 13 studies, Chyung et al. observed that many studies (paper and web based, with children and adults, in English and other language):

- Revealed response-order effects in self administered surveys, especially the primacy effect, associated with left-side selection bias, acquiescence bias, and satisficing.
- Showed more positive average scores from descending-ordered scales.

Recommendations:

- Present response scales in ascending order.
  - When a number line is used, lower and negative numbers should be on the left.
- When using descended order scales:
  - keep respondents motivated to complete items accurately,
  - present half items with descended-ordered scales and the other half with ascended-ordered scales,
  - assign half of participants with descended-ordered scales; half with ascended-ordered scales, and
  - present response options vertically rather than horizontally.

#### 2.3.4 Should surveys include negatively worded items?

In examining this question, Chyung et al. [Chyung et al., 2018c] made a distinction between (see Table 1 in the article):

- **Statement format** with a consistent response scale (e.g., strongly disagree to strongly agree).
- **Question format** with variable response scales that are tailored to individual survey questions.
  - A challenge with this format is the difficulty in calculating an average score of data obtained from multiple survey items.

The advent of negatively-worded items began with Rensis Likert in 1932. He was an American social psychologist who, in attempt to mitigate acquiescence/yea-saying biases, recommended designing one half of survey items to be associated with agreement and the other half with disagreement.

Although Likert recommended “straightforward statements,” incorporating negative words can become quickly complicated. Table 2 in the Chyung paper shows that there are four ways of wording survey statements:

**Reverse-coding**, which is necessary when including negatively worded items in a scale, assumes that agreeing to a positively worded statement and disagreeing to its negatively worded counterpart are the same. Tables 3 and 4 in the Chyung et al., manuscript [2018c] show how this assumption may be faulty. A review of the negatively-worded-item literature suggested the following:

- Scales with all positively worded items yielded greater accuracy when compared with all negatively worded items or mixed worded items.
- Scores on positively and negatively worded items are not the same (e.g., strongly disagreeing to a positively worded statement is different from strongly agreeing to a negatively worded statement)
- Positively worded items produce higher means than negatively worded items. This may be due to
  - carelessness and fatigue in reading items,
  - the cognitive processing of positive and negative items may be different.
- A *method factor* has shown itself where exploratory approaches to factor analysis have produced separate factors with the negatively worded (or otherwise ambiguous) items creating their own factor. This results in a threat to construct validity and reliability.

Chyung, Barkin, and Ramsey [2018c] noted that respondent performance declines approximately 12 minutes after starting a survey. It appears that respondents increasingly fail to notice negatively worded statements even when there are efforts to draw their attention to them via bolding, underlining, or capitalizing the negated element (e.g., **not**). Thus, when negatively worded items are used, they should probably be presented early in the protocol.

Chyung et al [2018c] also cautioned about a response set bias that can occur when using all positively worded items. They recommended making design choices that enhance bias-free and accurate responding based on the research design.

- For example, attributes to be measured in some constructs (e.g., depression, anxiety) are, themselves, negative and so a negatively worded item may be most clear and appropriate.
- The inclusion (and subsequent analysis) of negatively phrased items may help *detect* acquiescence bias.
- Table 5 in the Chyung et al [2018c] manuscript provides some guidelines that are more nuanced when negative items must be included. For example,
  - Ensure that negatively worded items are true polar opposites and symmetrical (so they can be analyzed with the positively worded items).
  - Group negative items together (and forewarn/format so they are recognized as such).
  - Administer the survey when respondents are not fatigued.
  - Analyze the effect of the negatively worded items.

## 2.4 Construct-specific guidance

Across disciplines and constructs, there may be localized guidance. One domain-specific example is *self-efficacy*. In this case, construct-specific guidance addresses both the (a) content of the items and (b) formatting of the scales. Regarding content, even though there are some *general self-efficacy scales* Bandura's original definition suggests that scales and their items should be task specific (i.e., career decision-making self-efficacy, math self-efficacy). Further, Bandura [2006], recommended the following for self-efficacy scales:

1. Phrase items as “can do” rather than “will do.”
2. Maintain consistency with the self-efficacy construct definition (e.g., domain specific, a focus on capability rather than self-worth).
3. Include items that reflect gradations of challenge.
4. Ask individuals to rate their current (as opposed to future) operative capabilities.
5. Use 100-point continuous scaling.

## 2.5 Surveying in the Online Environment

Nearly a decade ago, a survey of human subjects review boards suggested that 94% of the IRB applications reviewed involved online or Web-based surveys [Buchanan and Hvizdak, 2009]. Thus, it is important to understand the online environment. A first set of considerations involve data security, identity, and permission (implicit and explicit).

The **IP address** as well as **longitude/latitude** has been a contentious issue for a number of years [Buchanan and Hvizdak, 2009]. EU data protection laws consider IP addresses as personally identifiable data; in the U.S., IP addresses typically fall outside the definition of “personal information.” In Qualtrics, the default is to collect and download the IP address (the “anonymize response” option can prevent this data from being collected). On the one hand it is helpful to know geographically “from where” participants are responding; on the other, some consider its capture to be a violation of privacy. Relatedly, **paradata** and **metadata** are data such as typing speed, changed answers, response times, and time spent on the survey. For both geolocation and paradata/metadata, a strong consideration is **fully informed consent** [Schober and Conrad, 2007]. Is it ethical to capture this information without explicitly saying so? A best practice is to provide complete descriptions of what data is being collected and provide a rationale (in the IRB application, if not directly in the informed consent) for why it is necessary. Survey tools like Qualtrics have strong options for anonymizing data (i.e., permanently deleting identifying information). If this option is being used, this information should be included in the informed consent (and, perhaps, recruiting materials).

The specific **survey tool** being used should be evaluated. Buchanan and Hvizdak [2009] argued that until each tool is vetted and its privacy policies and data security policies are understood, we cannot be certain how security, consent, and privacy are operationalized within the individual tools. For example, it is possible that tool creators *could* gather respondent data and repurpose it for their own marketing, for sale to other researchers, and so forth.

Online and web-based protocols increase our reach geographically and cross-culturally. A first impression might be that the online environment increases **access** [Schober and Conrad, 2007]. We should probably think twice about this presumption. Consider the decades of psychological research

based on White, college-educated, males. Does the online environment create another strata of privileged research with technology that may not be accessible in terms of both internet/technology as well as capacity/fluency with the tool? Additionally, does the way that the survey is promoted result in invitations that only occur within certain segments of the internet? If so, the results may not be representative. On the other hand, what are the risks of not adopting new technologies before everyone has them. Another consideration is cultural and language translation. These issues are addressed more completely in the lesson on [invariance testing](#).

When paper/pencil measures were administered in face-to-face settings (individually or in auditoriums of students) there was some degree of a **standardized protocol**. This is lost when surveys are administered online. Further, we cannot guarantee *who* is taking the survey. Increasingly, when surveys are offered through fee-based programs like mTurk and Prolific, bots have been trained to take the surveys and receive the incentive. Survey programs like Qualtrics now offer additional packages to help with security and bot-prevention. It is probably also wise to add attention-check items (e.g., “This is an attention check. Please answer ‘3’”). Another option is to include a final question that asks the respondent to ensure the integrity of the response. An example last item might be, “To what degree is the statement below true of you: *I read each question/item and provided answers that were true for me.*” The scaling for this item was a 100 point slider with “Untrue,” “Neither true or untrue,” and “True”.

When respondents are remote, what happens if they have a **negative reaction to the survey**? In a face-to-face context, debriefings can occur and referrals can be made. IRB committees are likely to consider the degree to which surveys may be upsetting and require resources for referral and assistance in the case of an adverse event.

**Security of test items** might also be concerning. It is inappropriate to use proprietary items without the permission of its author. If the security of items is important (e.g., SAT/GRE, intelligence test items, inkblots) because they are central to administration, how can they be protected in the virtual environment?

Consequently, when students in our programs write doctoral dissertations they are to include the following in their Method section.

- Describe how informed consent will be obtained in the online environment.
- Describe the level of identification that is collected. If the claim of “anonymous” or “de-identified” indicate whether/not this includes capturing the IP address; some researchers believe that capturing a computer’s IP address threatens anonymity.
- Describe the steps to be taken to ensure that respondents met the inclusion/exclusion criteria of the study.
- Anticipate and describe how the online (e.g., uncontrolled, public, distractions) setting might affect responses.
- Particularly if the survey contained sensitive materials, describe how respondents might access resources for debriefing or referral.
- Identify the permissions (from original authors or copyright holders) granted to reformat and post (on the internet) existing surveys. If items are considered to be secure (e.g., those on the MMPI or WAIS), identify steps taken to protect them.

## 2.6 In my Surveys

Because there isn't empirical data on every decision that we make in survey construction, I thought it might be useful for me to address some of the decisions that I find myself making in the online surveys I use in my own research.

### 2.6.1 Demographics and Background Information

A core value that I hope to reflect in the *ReCentering Psych Stats* series is to promote socially and culturally responsive research. Correspondingly, the information we collect from participants should ensure that they feel that their identities are authentically reflected in the survey. Naively, when I first considered how to capture race/ethnicity in my surveys, I looked to the categories used in the U.S. Census. Immediately, I learned that this is problematic. Rather than elaborating here, I invite you to take a listen to NPR's [Code Switch](#) podcast. Two of the episodes review how the assessment of race and ethnicity has evolved and explain why it is problematic: [Census Watch 2020](#) and [The U.S. Census and Our Sense of Us](#). As made clear in the Code Switch podcasts, the assessment of race and ethnicity in the U.S. Census *erases* people when their identities are not included.

My last few surveys have captured race/ethnicity data differently. Each time, I engage in several practices that (I hope) will continue to shape the item in a socially and culturally responsive way. Systematically, I:

- conduct a quick internet search to see if there is an emerging best practice (even though I may have also searched weeks or months prior),
- consider who the intended research population is in relationship to the topic of investigation,
- look to recently published, similar, research to see what other researchers are doing, and
- ask for a collegial, formative review from individuals who hold marginalized identities, whose data will be requested in the survey.

When I engage in research, I try to balance the need to quantify (with discrete categories) who is participating in the survey and inviting respondents to state (in their own words) their identity. This is consistent with my view that variables like race, ethnicity, and gender identity are socially constructed. In addition to this particular worldview, Parent [2013] has suggested that the worst possible kind of missing data pattern (MNAR – missing not at random) may be caused when items are *unanswerable* to particular person. Therefore, it is essential that all individuals recognize themselves in the items that assess demographic variables.

A recent survey of mine was directed toward community members (including students, alumni, staff, faculty) of my predominantly White, liberal arts, Christian institution. After reviewing recently published research articles and consulting with a handful of individuals, I chose to include the following categories – *each with a text write-in box* so that individuals could select the category(ies) that fit best and have the opportunity to refine it(them). I am excited to review this data because such responses may inform my next survey. The categories included:

- Asian or Pacific Islander
- Black or African American

- Hispanic or Latino
- Native American or Alaskan Native
- White or Caucasian
- Biracial or multiracial
- An international/global identity that does not fit in the U.S. categorization of race/ethnicity
- A race/ethnicity not listed above

Respondents could select multiple categories. Additionally, they could write in an identity that better aligned with their self-understanding.

RaceEthnicity ★

Which of the following best describes you?

*Owing to the limitations of surveys, we have provided categories typically used in survey construction. Please select the category(ies) that is(are) closest to the manner in which you identify. You are welcome to refine the response in the text box.*

Asian or Pacific Islander  
 Black or African American  
 Hispanic or Latino  
 Native American or Alaskan Native  
 White or Caucasian  
 Biracial or multiracial  
 An international/global identity that does not fit in the U.S. categorization of race/ethnicity  
 A race/ethnicity not listed above

Figure 2.1: Image of a survey item inquiring about race/ethnicity from a survey. Each option has an option for the respondent to clarify.

The option to select multiple boxes results in some extra coding when preparing the data for analysis. I am taking approach that we will *listen* to the data and decide, based on the results, how to report the findings in a way that will efficiently fit into an APA style empirical paper and honor the respondents.

The population of interest for this particular study are those who are engaged in protest activities regarding hiring practices and policies that result in discrimination to members of the LGBTQIA+ community. This means that questions of gender identity, pronouns, and relationship to the LGBTQIA+ community are important to the research, and need to be asked sensitively and with great security of the data.

Regarding gender identity, I used a similar approach, allowing individuals to select multiple categories and offering write-in boxes for each. The categories included:

- female
- male
- nonbinary
- trans woman
- trans man
- another identity (with a write-in box)
- prefer not to say

Additionally, I invited individuals to identify their pronouns. Again, write-in boxes were offered with each option.

- they/them/theirs
- she/her/hers
- he/him/his
- they/she/he
- neo pronouns (e.g., xe/xem/xyr, ze/hir/hirs, ey/em/eir)
- another identity (with a write-in box)

Finally, we wanted individuals to indicate their relationship to the LGBTQIA+ community. We asked them to select all that apply. Only the “something else” box had a write-in option:

- Member
- Exploring/Questioning
- Ally
- Not related
- Something else (with a write-in box)

I expect that my future surveys may inquire about these variables differently. If you have found a different way to ask, please consider e-mailing me. I would love to provide different options and give credit to contributors.

### 2.6.2 Survey Order

Historically, demographic information has been first or last in the survey. Although some research has reported no differences in response rates when demographic and sensitive data are at the beginning or end [Krathwohl, 2009, Rowley, 2014], I am inclined to open the survey with questionnaire items that are closely related to the topic listed on the recruitment materials and end the survey with the demographic information. Why? It makes sense to me that if someone has responded positively to the survey topic, they expect to answer questions (right away) about that topic.

In between that opening survey and closing demographic items, I consider if there are any *order effects* that would engage in undue *priming* of responses. If there are such concerns, I think through the order to minimize these biasing effects. If there are no such concerns, I put my surveys in blocks and then ask my survey program to randomly present the blocks. This serves two purposes:

- counterbalancing possible order effects, and
- distributing missingness for individuals who do not complete the survey.

### 2.6.3 Forced Responses

Programs like Qualtrics are able to engage in a variety of *content validation* procedures. If these are in place, they may require the person to enter a properly formatted response (e.g., phone number, e-mail address, numerical response between 0 and 100) before responding. These are

extremely helpful tools in collecting data that will be closest-to-being-ready-for-analysis. These same procedures can *force* or *request* a response.

*Requiring* a response is tempting. However, doing so violates IRB requirements that allow a person to skip or “quit at any time without penalty.” They may also anger a person such that they stop responding. Some researchers get around this by *requiring* the response, but including a “Not applicable” or “Prefer to not answer” column. Because I worry that (a) the respondent may confuse that option with one extreme of the scale and/or (b) my research team and I will forget to code it as missing data, I prefer the *request* response alternative.

In Qualtrics in particular, I turn on the “Request response” feature for each of the questions. If an item is skipped, a simple warning is displayed that invites the respondent to review the page of answers to see if they would like to answer the question. If not, they can simply move forward.

## 2.7 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. With each of these options I encourage you to:

This is a two-part lesson on questionnaire construction. After the [second lesson](#), a detailed suggestion for practice will be provided that lists criteria for creating and piloting a survey of your own.

# Chapter 3

## Be a QualTRIXter

### [Screencasted Lecture Link](#)

The focus of this lecture is on the technical and mechanical tools available in Qualtrics (and likely other survey platforms) to increase the effectiveness of your survey.

### 3.1 Navigating this Lesson

This lecture is just under one hour. Plan for another 30 minutes for *intRavenous qualtRics* practice.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

#### 3.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Utilize basic Qualtrics tools (e.g., question type, use of headers) so that surveys are present materials clearly to the respondent.
- Incorporate more advanced tools (e.g., display logic, randomization) that may increase the respondent's ability to complete the survey and provide accurate responses.
- Provide a rationale for survey options that protect (or possibly reveal) an individual's identity.

#### 3.1.2 Planning for Practice

This is the second of a two-part lesson on questionnaire construction. At the end of this lesson is a detailed suggestion for practice that lists criteria for creating and piloting a survey of your own. There are four essential criteria for your survey:

- Adhere to the evidence-based practices identified in the lesson on [questionnaire construction](#).

- Utilize four techniques (in the context of Qualtrics, I term these *qualTRIXter skills*) that increase the flow, effectiveness, and appearance of your survey.
- Pilot and consider feedback provided by those who took the survey.
- Import the data into the R environment.

### 3.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the tutorials available at the [Qualtrics support site](#). I have tried to link them throughout the presentation. It is likely they could change at any time and/or they might not work on your particular browser.

### 3.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(qualtrics)){install.packages('qualtrics')}
```

## 3.2 Research Vignette

I will demonstrate the qual”TRIX” by using a Qualtrics account hosted at Seattle Pacific University. The only surveys in this account are for the *Recentering Psych Stats* chapters and lessons. All surveys are designed to not capture personally identifying information and not collecting IP addresses nor longitude/latitude. I use this survey in several lessons in this OER. If you haven’t taken the survey yet, [I invite you to do so, now](#).

As a teaching activity for the ReCentering Psych Stats OER, the topic of the survey was selected to be consistent with the overall theme of OER. Specifically, the purpose of this study is to understand the campus climate for students whose identities make them vulnerable to bias and discrimination. These include students who are Black, non-Black students of color, LGBTQ+ students, international students, and students with disabilities.

After consulting with a diverse group of stakeholders and subject matter experts (and revising the response options numerous times) I have attempted to center anti-Black racism in the U.S. [Mosley et al., 2021, 2020, Singh, 2020]. In fact, the display logic does not present the race items when the course is offered outside the U.S. There are only five options for race: *biracial/multiracial*, *Black*, *non-Black person(s) of color*, *White*, and *I did not notice* (intended to capture a color-blind response). One unintended negative consequence of this design is that the response options could contribute to *colorism* [Adames et al., 2021, Capielo Rosario et al., 2019]. Another possibility is that the limited options may erase, or make invisible, other identities. At the time that I wrote up the first description of this survey, the murder of six Asian American women in Atlanta had just occurred. The Center for the Study of Hate and Extremism has documented that while overall hate crimes dropped by 7% in 2020, anti-Asian hate crimes reported to the police in America’s largest cities increased by 149% [noa, a]. These incidents have occurred not only in cities, but in

our neighborhoods and on our campus [Kim, 2021b,a, noa, b]. While this survey is intended to assess campus climate as a function of race, it unfortunately does not distinguish between many identities that experience marginalization.

Although the dataset should provide the opportunity to test a number of statistical models, one working hypothesis that framed the study is that there will be a greater sense of belonging and less bias and discrimination when there is similar representation (of identities that are often marginalized) in the instructional faculty and student body. Termed, “structural diversity” [Lewis and Shah, 2019] this is likely an oversimplification. In fact, an increase in diverse representation without attention to interacting factors can increase hostility on campus [Hurtado, 2007]. Thus, the task of rating of a single course relates to the larger campus along the dimensions of belonging and bias/discrimination. For example, if a single class has higher ratings on issues of inclusivity, diversity, and respect, we would expect that sentiment to be echoed in the broader institution.

The survey design has notable limitations You will likely notice that we ask about demographic characteristics of the instructional staff and classmates in the course rated, but we do not ask about the demographic characteristics of the respondent. In making this decision, we likely lose important information. For example, Iacobino and James [2016] have noted that White students perceive campus more favorably than Black student counterparts.

The decision to not collect demographic details about the respondent was about protecting their (your) identity. As you will see, you have the opportunity to download and analyze the data. If a faculty member asked an entire class to take the survey, the datestamp and a handful of demographic identifiers could very likely identify a student. In certain circumstances, this might be risky in that private information (i.e., gender nonconformity, disclosure of a disability) along with course evaluation data and a date stamp could identify the respondent.

Further, the items that ask respondents to *guess* the identities of the instructional staff and classmates are limited, and contrary to best practices in survey construction that recommend providing the option of a “write-in” a response.

In parallel, the items asking respondents to identify characteristics of the instructional staff along dimensions of gender, international status, and disability are “large buckets” and do not include “write-in” options. Similarly, there was no intent to cause harm by erasing or making invisible individuals whose identities are better defined by different descriptors. Further, no write-in items were allowed. This was also intentional to prevent potential harm caused by people who could leave inappropriate, racist, or otherwise harmful comments.

As I review Qualtrics essentials and trix, I will their use (if used) in the ReCentering Psych Stats survey.

### 3.3 Qualtrics Essentials

Qualtrics is a powerful program and I find that many of the surveys we distribute don’t capitalize on the features Qualtrics has to offer. Qualtrics has detailed tutorials and instructions that are well worth the investment of a weekend to review them.

In this lecture I will point you to the elements that I think are critical to constructing online surveys. Because Qualtrics tutorials are (a) clear and thorough and (b) frequently updated, I will (a) point you to the tutorials that are available at the time of this lecture prep, (b) tell you why I think they are appropriate, and (c) show you how we have used them in some of our own surveys.

Even if you think you know what you are doing, start here (and then always take the time to “look around” at all the options on each window):

**Survey Basic Overview:** Qualtrics’ [Survey Basic Overview](#) tutorial is a great place to start. From there, you can follow all kinds of leads, looking for things you want to do with your survey – and getting ideas for what will improve it.

**Blocks** are the basic organizational tool in Qualtrics surveys. Blocks have two purposes: (a) grouping items shown on “one page,” and (b) specifying ordering and/or random selection/presentation in the survey flow.

**Question types:** Take time to look at all the options. You might be surprised to learn that there is a better choice than you might have imagined.

Let’s take a look at super basic/helpful question types:

- **Text/graphic:** These are the types you should use for providing information (e.g., informed consent) to the participants or displaying a logo or graphic stimulus.
- **Matrix table:** The matrix table is a more efficient way to use the Likert-style items (than multiple choice). There is some controversy about whether not to use matrix tables vs. multiple choice dropdowns. As both a survey developer and a respondent, I prefer the matrix table.
  - Make sure to select a reasonable amount of header repetitions. This allows the respondent the maximum opportunity to see the column descriptors (and avoid guessing/remembering) while they are responding.
- **Slider :** The slider is designed for obtaining truly continuous data on a 1 to 100 scale. This range can be adapted to any interval you choose and you can add anchors to the scale. If the scale you are using is already published, and has not been psychometrically evaluated for slider use, you should probably stick with the format recommended in the publication. But if you are writing test items, consider this option.
- **Text Entry Questions:** Text boxes have multiple options for answer length.
- **Validation:** Content validation allows the user to permit certain types of information and specify their formats (e.g., numbers, e-mail addresses, dates). There is art to balancing between being overly restricting and ensuring that the data is entered in the most clear and consistent way possible with honoring the uniqueness of each respondent. Another validation option I frequently use is one that asks individuals if they intended to leave something blank. This is tool that helps prevent missingness without forcing an individual to respond to an item that (a) might not be clear to them, (b) might not be appropriate for them, and/or (c) might result in an answer that is untrue for their unique circumstance.

## 3.4 Qual-TRIX

**Collaborating** with other Qualtrics users in your institution is easy! Scroll down to “Collaborating Inside Your Organization” and follow the instructions for adding individuals to your survey (you must “own” the survey; your collaborators will not be able to add others).

The ability to **schedule survey distributions** is like having your very own assistant! If you have a roster (contact list) you can schedule distributions, reminders, and thank yous. Qualtrics will keep track of who responds and send reminders to the non-responders. Here are resources for

- [E-mail overview](#)
- [E-mail distribution management](#)
- [Directories](#)

**Personalizing** invitations and surveys. **Piped text** is a way to personalize invitations and/or “carry forward” prior responses into new questions.

**Randomization of blocks** (or a subset of blocks) can be used for several purposes such as: (a) using random selection to display one or more blocks to respondents – as in a random clinical trial, (b) to randomly display a percentage of blocks or items to shorten the survey in a planned missing design, and (c) randomly display some or all of the blocks of the survey to all respondents so that when respondents experience test fatigue, when they quit responding, “the last items/surveys” aren’t always the same ones. This functions to distribute missingness across surveys.

**Randomization of items** within a block can be used for similar purposes. You can also use this to display only some of the items (e.g., planned missingness).

**File upload from respondents** is an additional package that requires the institution to pay a higher fee. If available, this allows respondents to upload some sort of file (photo, powerpoint, .pdf). We use it for poster contests at professional contests (where students upload their poster for online judging in advance of the conference). A colleague of mine uses this function to collect application elements (i.e., resumes, cover letters, reference letters) to a fellowship program.

- As researchers, we can also upload files (e.g., hardcopy of informed consent, documents to be reviewed) for use by the respondent.

**Display, Skip, and/or Branch Logic** can be used to help display to respondents *only* the items that pertain to them. There are multiple approaches to doing this. Using a display logic approach may feel a bit *backward* where the logic is applied *from* the landing spot. We did this extensively in a study that involved two language versions and three age options.

Two other approaches for these issues are [skip logic](#) and [branch logic](#)

### 3.5 Even moRe, particularly relevant to iRb

We can use Qualtrics tools for purposes beyond collecting and downloading data. These tools are especially useful when I think about IRB applications and ethics related to data collection.

**Exporting to Word:** Helpful for your IRB application (and perhaps in a cloud so that a team can use track changes to edit), it is super simple to export the survey to Microsoft Word. Additionally, you can specify options for including question numbers, recode values, logic, and so forth. This works well to create a codebook for your research team.

**Anonymizing responses:** Another step toward an anonymous response is to withhold the IP address, latitude/longitude, and any contact information (e.g., e-mail, name) that you may have uploaded in an e-mail distribution directory. This is accomplished in the Survey Options menu. Do be careful – while anonymizing responses is an ethical, best practice, the deleted information cannot be recovered.

**Prevent ballot box stuffing:** Want to make sure that respondents only answer once? In the same Survey Options window, you can prevent ballot box stuffing. This is helpful when surveys are distributed with an *anonymous link*. The tool prevents more than one survey from the same IP address.

Other security options include

- Password protection
- HTTP Referer verification

Look also at:

- **Progress bar** to provide participants hope (or despair) for “how much longer.”
- **Survey termination** to connect custom endings and thank-you notes.
- **Partial completion** to specify how long the respondent has to complete the survey (after opening it) and whether it is recorded or deleted if it is not completed.
  - Related to this, back on the *Data & Analysis* tab, you can see both the numbers of **recorded responses** and **responses in progress**. You also have options to manually determine how you want to include/exclude the responses in progress.
  - Failure of the respondent to click the final “->” submit and progress symbol is often the reason that surveys that are > 90% complete aren’t counted as “complete.” What to do? Options: (a) don’t say “Thanks and goodbye” on a page that has any items, and (b) provide instructions to look for the “->” symbol to continue.

Finally, **PREVIEW PREVIEW PREVIEW!** There is no better way check your work than with previews.

## 3.6 intRavenous Qualtrics

Access credentials for the institutional account, individual user’s account, and survey are essential for getting the survey items and/or results to export into R. The Qualtrics website provides a tutorial for generating an [API token](#).

We need two pieces of information: the **root\_url** and an **API token**.

- Log into your respective qualtrics.com account.
- Select Account Settings
- Choose “Qualtrics IDs” from the user name dropdown

We need the **root\_url**. This is the first part of the web address for the Qualtrics account. For our institution it is: spupsych.az1.qualtrics.com

The API token is in the box labeled, “API.” If it is empty, select, “Generate Token.” If you do not have this option, locate the *brand administrator* for your Qualtrics account. They will need to set up your account so that you have API privileges.

*BE CAREFUL WITH THE API TOKEN* This is the key to your Qualtrics accounts. If you leave it in an .rmd file that you forward to someone else, this key and the base URL gives access to every survey in your account. If you share it, you could be releasing survey data to others that would violate confidentiality promises in an IRB application.

If you mistakenly give out your API token you can generate a new one within your Qualtrics account and re-protect all its contents.

You do need to change the API key/token if you want to download data from a different Qualtrics account. If your list of surveys generates the wrong set of surveys, restart R, make sure you have the correct API token and try again.

```
# only have to run this ONCE to draw from the same Qualtrics
# account...but will need to get different #token if you are changing
# between accounts.
qualtrics::qualtrics_api_credentials(api_key = "oEwd9qu9xJ0f3RoE9iiCZKSs2sfNuSbvy8LnFYxo",
  base_url = "spupsych.az1.qualtrics.com", overwrite = TRUE, install = TRUE)
readRenvironment("~/.Renviron")
```

*all\_surveys()* generates a dataframe containing information about all the surveys stored on your Qualtrics account.

```
surveys <- qualtrics::all_surveys()
# View this as an object (found in the right: Environment). Get
# survey id # for the next command If this is showing you the WRONG
# list of surveys, you are pulling from the wrong Qualtrics account
# (i.e., maybe this one instead of your own). Go back and change your
# API token (it saves your old one). Changing the API likely requires
# a restart of R.
surveys
```

To retrieve the survey, use the *fetch\_survey()* function.

```
# obtained with the survey ID
#'surveyID' should be the ID from above
#'verbose' prints messages to the R console
#"label", when TRUE, imports data as text responses; if FALSE prints the data as numerical responses
#"convert", when TRUE, attempts to convert certain question types to the 'proper' data type in the output
#"force_request", when TRUE, always downloads the survey from the API instead of from a temporary URL
#"import_id", when TRUE includes the unique Qualtrics-assigned ID;
# since I have provided labels, I want false
```

```
# Out of the blue, I started getting an error, that R couldn't find
# function 'fetch_survey.' After trying a million things, adding
# qualRics:: to the front of it solved the problem
QTRX_df <- qualRics::fetch_survey(surveyID = "SV_b2cClqA1LGQ6nLU", time_zone = NULL,
  verbose = FALSE, label = FALSE, convert = FALSE, force_request = TRUE,
  import_id = FALSE)

# useLocalTime = TRUE,
```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do).

```
# write the simulated data as a .csv write.table(QTRX_df,
# file='QTRX_df.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file QTRX_df <- read.csv
# ('QTRX_df.csv', header = TRUE)

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(QTRX_df, 'QTRX_df.rds') bring back the simulated dat
# from an .rds file QTRX_df <- readRDS('QTRX_df.rds')
```

### 3.6.1 The Codebook

In order to prepare data from a survey, it is critical to know about its content, scoring directions for scales/subscales, and its design. As I demonstrated above, we can export a **codebook**, that is, a Word (or PDF) version of the survey with all the coding. In Qualtrics the protocol is: Survey/Tools/ImportExport/Export Survey to Word. Then select all the options you want (especially “Show Coded Values”). A tutorial provided by Qualtrics can be found [here](#). This same process can be used to print the PDF example I used above.

I recommend providing custom variable names and recode values directly in Qualtrics before exporting them into R (and before exporting the codebook). A Qualtrics tutorial for this is provided [here](#). In general, consider these qualities when creating variable names:

- Brevity: historically, SPSS variable names could be a maximum of 8 characters.
- Intuitive: although variables can be renamed in R (e.g., for use in charts and tables), it is helpful when the name imported from Qualtrics provides some indication of what the variable is.
- Systematic: start items in a scale with the same stem, followed by the item number – ITEM1, ITEM2, ITEM3.
- Do not include special characters or spaces in variable names; this is problematic for R.
- Do not start variable names with numerals; this is problematic for R.

More complete information about data preparation is covered in chapters in the [ReCentering Psych Stats: Multivariate Modeling](#) text.

### 3.6.2 Using data from an exported Qualtrics .csv file

It is also possible to download the Qualtrics data in a variety of formats (e.g., CSV, Excel, SPSS). Since my R and Qualtrics history began by using files with the CSV extension (think “Excel” lite), that is my preference.

In Qualtrics, these are the steps to download the data: Projects/YOURsurvey/Data & Analysis/Export & Import/Export data/CSV/Use numeric values. In order to import this data into R, it is critical that to save this file in the same folder as the .rmd file that you will use with the data.

R is sensitive to characters used in filenames. As downloaded, my Qualtrics .csv file had a long name with spaces and symbols that are not allowed. Therefore, I gave it a simple, sensible, filename, “ReC\_Download210319.csv”. An idiosyncrasy of mine is to datestamp filenames. I use two-digit representations of the year, month, and date so that if the letters preceding the date are the same, the files would alphabetize automatically.

```
# QTRX_csv <- qualtrics::read_survey('ReC_Download210319.csv',
# strip_html = TRUE, import_id = FALSE, time_zone=NULL, legacy =
# FALSE)
```

Although minor tweaking may be required, the same script above should be applicable to this version of the data.

### 3.6.3 Tweaking Data Format

Two general approaches:

1. Inside Qualtrics: Use the [recode values](#) option (found in the item’s gearbox, to the left of the block) to specify variable names and recode values. These should be preserved on the download.
2. In the R script: In another lecture I demonstrate how to change the formats of data (character, string), selecting only the variables in which we are interested (e.g., excluding the meta-data), and renaming variables sensibly.

Both work! You can choose your preference. When you are working with a team, map out an explicit process with your collaborators.

## 3.7 Practice Problems

The suggestion for practice is to develop a questionnaire, format it, pilot it, and download it. Essentially you will be

- Formatting a survey on Qualtrics using all the best practices identified in the lecture. These include:

- Having an introductory statement (to include statement of confidentiality), directions for each sub-survey (if more than one), and closing statement.
- Selecting the most appropriate question type for the items. For example, matrix instead of multiple choice.
- Within the question type, using the appropriate options for proper formatting (e.g., the anchors in a matrix should be topically consistent and equal-interval).
- The survey should include minimum of 3 of the qualTRIXter skills (identified in lecture). Choose from:
  - establishing collaboration
  - scheduling e-mail distribution and follow-up
  - personalizing the survey in some way
  - randomization of blocks or items
  - integrating display, skip, or branch logic (e.g., having males and females take a different route)
  - exporting the survey to Word
  - recoding variables in the item controls
  - anonymizing the responses
  - preventing ballot box stuffing
  - including a progress bar
  - creating a custom ending, e-mail, or thank-you note
  - something else that YOU discovered that isn't in the lecture
- Piloting it, getting their feedback, and identifying what problems are (and how you might fix them).
  - with 3 folks from your research team, cohort, or this class
  - with 3 additional folks who aren't quite as “research savvy”
  - collect their feedback (ideally in a text-item directly on the survey itself) and write a brief summary (3 paragraphs max) of their impressions and how you might improve the survey
- Import the Qualtrics data directly R
  - preferably, directly from Qualtrics with the API token, base URL, and survey ID
  - alternatively (for the same # of points) from the exported CSV file *via the qualtrics package* (required)

Assignment Component	Points Possible	Points Earned
1. Qualtrics survey best practices	5	
2. QualTRIXter skills (at least 3)	5	
3. Minimum of 6 pilot respondents	5	
4. Summary of pilot feedback	5	
5. Import of Qualtrics data into R	5	
6. Explanation to grader	5	
<b>Totals</b>	<b>20</b>	

# Chapter 4

## Psychometric Validity: Basic Concepts

### [Screencasted Lecture Link](#)

The focus of this lecture is to provide an introduction to validity. This includes understanding some of the concerns of validity, different aspects of validity, and factors as they affect validity coefficients.

### 4.1 Navigating this Lesson

There is just over one hour of lecture. If you work through the materials with me it would be plan for an additional hour.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

#### 4.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Distinguish between different types of validity based on short descriptions.
- Compute and interpret validity coefficients.
- Evaluate the incremental validity of an instrument-of-interest.
- Define and interpret the standard error of estimate.
- Develop a rationale that defends importance of establishing the validity of a measuring instrument.

### 4.1.2 Planning for Practice

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. With each of these options I encourage you to interpret examine aspects of the construct validity through the creation and interpretation of validity coefficients. Ideally, you will examine both convergent/discriminant as well as incremental validity.

### 4.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

- Jhangiani, R. S., Chiang, I.-C. A., Cuttler, C., & Leighton, D. C. (2019). Reliability and Validity. In *Research Methods in Psychology*. <https://doi.org/10.17605/OSF.IO/HF7DQ>
- Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
  - In this manuscript, Clark and Watson (1995) create a beautiful blend of theoretical issues and practical suggestions for creating measures that evidence construct validity. From the practical perspective, the authors first guide potential scale constructors through the literature review and creating an item pool (including tips on writing items). The authors address structural validity by first beginning with strategies for constructing the test. In this section, the authors revisit the issue of dimensionality (i.e., alpha vs. factor analysis). Finally, the authors look at initial data collection (addressing sample size) and psychometric evaluation.

### 4.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(tidyverse)){install.packages('tidyverse')}
# if(!require(MASS)){install.packages('MASS')}
# if(!require(psych)){install.packages('psych')}
```

## 4.2 Research Vignette

This lesson provides descriptions of numerous pathways for establishing an instrument's validity. In fact, best practices involving numerous demonstrations of validity. Across several lessons, we will rework several of the correlational analyses reported in the research vignette. For this lesson in particular, the research vignette allows demonstrations of convergent/discriminant validity and incremental validity.

The research vignette for this lesson is the development and psychometric evaluation of the Perceptions of the LGBTQ College Campus Climate Scale [Szymanski and Bissonette, 2020]. The scale is six items with responses rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores indicate more negative perceptions of the LGBTQ campus climate. Szymanski and Bissonette have suggested that the psychometric evaluation supports using the scale in its entirety or as subscales composed of the following items:

- College Response to LGBTQ students:
  - My university/college is cold and uncaring toward LGBTQ students.
  - My university/college is unresponsive to the needs of LGBTQ students.
  - My university/college provides a supportive environment for LGBTQ students.  
[un]supportive; must be reverse-scored
- LGBTQ Stigma:
  - Negative attitudes toward LGBTQ persons are openly expressed on my university/college campus.
  - Heterosexism, homophobia, biphobia, transphobia, and cissexism are visible on my university/college campus.
  - LGBTQ students are harassed on my university/college campus.

A preprint of the article is available at ResearchGate.

Because data is collected at the item level (and I want this resource to be as practical as possible), I have simulated the data for each of the scales utilized in the research vignette at the item level. Simulating the data involved using factor loadings, means, standard deviations, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below re-scales the scores into the range of the Likert-type scaling and rounds them to whole values.

Five additional scales were reported in the Szymanski and Bissonette article [2020]. Unfortunately, I could not locate factor loadings for all of them; and in two cases, I used estimates from a more recent psychometric analysis. When the individual item and their factor loadings were known, I assigned names based on item content (e.g., “lo\_energy”) rather than using item numbers (e.g., “PHQ4”). When I am doing psychometric analyses, I prefer item-level names so that I can quickly see (without having to look up the item content) how the items are behaving. While the focus of this series of chapters is on the LGBTQ Campus Climate scale, this simulated data might be useful to you in one or more of the suggestions for practice (e.g., examining the psychometric characteristics of one or the other scales). The scales, their original citation, and information about how I simulated data for each are listed below.

- **Sexual Orientation-Based Campus Victimization Scale** [Herek, 1993] is a 9-item item scale with Likert scaling ranging from 0 (*never*) to 3 (*two or more times*). Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.
- **College Satisfaction Scale** [Helm et al., 1998] is a 5-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores represent greater college satisfaction. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.

- **Institutional and Goals Commitment** [Pascarella and Terenzini, 1980] is a 6-item subscale from a 35-item measure assessing academic/social integration and institutional/goal commitment (5 subscales total). The measure had with Likert scaling ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Higher scores on the institutional and goals commitment subscale indicate greater intentions to persist in college. Data were simulated using factor loadings in the source article.
- **GAD-7** [Spitzer et al., 2006] is a 7-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate more anxiety. I simulated data by estimating factor loadings from Brattmyr et al. [2022].
- **PHQ-9** [Kroenke et al., 2001] is a 9-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate higher levels of depression. I simulated data by estimating factor loadings from Brattmyr et al. [2022].

#Entering the intercorrelations, means, and standard deviations from the journal article

```
Szymanski_generating_model <- '
  #measurement model
  CollegeResponse  =~ .88*cold + .73*unresponsive + .73*supportive
  Stigma =~ .86*negative + .76*heterosexism + .71*harassed
  Victimization =~ .8*Vic1 + .8*Vic2 + .8*Vic3 + .8*Vic4 + .8*Vic5 + .8*Vic6 + .8*Vic7 + .8*Vic8 + .8*Vic9
  CollSat =~ .8*Sat1 + .8*Sat2 + .8*Sat3 + .8*Sat4 + .8*Sat5
  Persistence =~ .69*graduation_importance + .63*right_decision + .62*will_register + .58*will_study + .58*will_research
  Anxiety =~ .851*nervous + .887*worry_control + .894*much_worry + 674*cant_relax + .484*anxious
  Depression =~ .798*anhedonia + .425*down + .591*sleep + .913*lo_energy + .441*appetite

  #Means
  CollegeResponse ~ 2.71*1
  Stigma ~3.61*1
  Victimization ~ 0.11*1
  CollSat ~ 5.61*1
  Persistence ~ 4.41*1
  Anxiety ~ 1.45*1
  Depression ~1.29*1

  #Correlations
  CollegeResponse ~~ .58*Stigma
  CollegeResponse ~~ -.25*Victimization
  CollegeResponse ~~ -.59*CollSat
  CollegeResponse ~~ -.29*Persistence
  CollegeResponse ~~ .17*Anxiety
  CollegeResponse ~~ .18*Depression

  Stigma ~~ .37*Victimization
  Stigma ~~ -.41*CollSat
  Stigma ~~ -.19*Persistence
  Stigma ~~ .27*Anxiety
```

```

Stigma ~~ .24*Depression

Victimization ~~ -.22*CollSat
Victimization ~~ -.04*Persistence
Victimization ~~ .23*Anxiety
Victimization ~~ .21*Depression

CollSat ~~ .53*Persistence
CollSat ~~ -.29*Anxiety
CollSat ~~ -.32*Depression

Persistence ~~ -.22*Anxiety
Persistence ~~ -.26*Depression

Anxiety ~~ .76*Depression
'

set.seed(240218)
dfSzy <- lavaan::simulateData(model = Szymanski_generating_model,
                               model.type = "sem",
                               meanstructure = T,
                               sample.nobs=646,
                               standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(dfSzy))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 6 are the Perceptions of LGBTQ Campus Climate Scale
#Rows 7 thru 15 are the Sexual Orientation-Based Campus Victimization Scale
#Rows 16 thru 20 are the College Satisfaction Scale
#Rows 21 thru 26 are the Institutional and Goals Commitment Scale
#Rows 27 thru 33 are the GAD7
#Rows 34 thru 42 are the PHQ9

for(i in 1:ncol(dfSzy)){
  if(i >= 1 & i <= 6){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
  }
  if(i >= 7 & i <= 15){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
  }
  if(i >= 16 & i <= 20){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
  }
  if(i >= 21 & i <= 26){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 5))
  }
}

```

```

        }
if(i >= 27 & i <= 33){
  dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
}
if(i >= 34 & i <= 42){
  dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
}
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
dfSzy <- dfSzy %>% round(0)

#quick check of my work
#psych::describe(dfSzy)

#Reversing the supportive item on the Perceptions of LGBTQ Campus Climate Scale so that the exercises
dfSzy <- dfSzy %>%
  dplyr::mutate(supportiveNR = 8 - supportive)

#Reversing three items on the Institutional and Goals Commitments scale so that the exercises
dfSzy <- dfSzy %>%
  dplyr::mutate(not_graduateNR = 8 - not_graduate)%>%
  dplyr::mutate(undecidedNR = 8 - undecided)%>%
  dplyr::mutate(grades_unimportantNR = 8 - grades_unimportant)

dfSzy <- dplyr::select(dfSzy, -c(supportive, not_graduate, undecided, grades_unimportant))

```

The optional script below will let you save the simulated data to your computing environment as either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(dfSzy, 'SzyDF.rds') bring back the simulated dat from
# an .rds file dfSzy <- readRDS('SzyDF.rds')

```

```

# write the simulated data as a .csv write.table(dfSzy,
# file='SzyDF.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file
dfSzy <- read.csv("SzyDF.csv", header = TRUE)

```

As we move into the lecture, allow me to provide a content advisory. Individuals who hold LGBTQIA+ identities are frequently the recipients of discrimination and harassment. If you are curious about why these items are considered to be stigmatizing or non-responsive, please do not

ask a member of the LGBTQIA+ community to explain it to you; it is not their job to educate others on discrimination, harassment, and microaggressions. Rather, please read the article in its entirety. Additionally, resources such as [The Trevor Project](#), [GLSEN](#), and [Campus Pride](#) are credible sources of information for learning more.

## 4.3 Fundamentals of Validity

**Validity** (the classic definition) is the ability of a test to measure what it purports to measure. Supporting that definition are these notions:

- Validity is extent of matching, congruence, or “goodness of fit” between the operational definition and concept it is supposed to measure.
- An instrument is said to be valid if it taps the concept it claims to measure.
- Validity is the appropriateness of the interpretation of the results of an assessment procedure for a given group of individuals, not to the procedure itself.
- Validity is a matter of degree; it does not exist on an all-or-none basis.
- Validity is always specific to some particular use or interpretation.
- Validity is a unitary concept.
- Validity involves an overall evaluative judgment.

Over the years (and, perhaps within each construct), validity has somewhat of an *evolutionary* path from a focus on content, to prediction, to theory and hypothesis testing.

When the focus is on **content**, we are concerned with the:

- Assessment of what individuals had learned in specific content areas.
- Relevance of its content (i.e., we compare the content to the content domain).

When the focus is on **prediction**, we are concerned with:

- How different persons respond in a given situation (now or later).
- The correlation coefficient between test scores (predictor) and the assessment of a criterion (performance in a situation)

A focus on **theory and hypothesis testing** adds:

- A strengthened theoretical orientation.
- A close linkage between psychological theory and verification through empirical and experimental hypothesis testing.
- An emphasis on constructs in describing and understanding human behavior.

**Constructs** are broad categories, derived from the common features shared by directly observable behavioral variables. They are theoretical entities and not directly observable. **Construct validity** is at the heart of psychometric evaluation. We define **construct validity** as the fundamental and all-inclusive validity concept, insofar as it specifies what the test measures. Content and predictive validation procedures are among the many sources of information that contribute to the understanding of the constructs assessed by a test.

## 4.4 Validity Criteria

We have just stated that validity is an overall, evaluative judgment. Within that umbrella are different criteria by which we judge the validity of a measure. We casually refer to them as *types*, but each speaks to that unitary concept.

### 4.4.1 Content Validity

Content validity is concerned with the representativeness of the domain being assessed. Content validation procedures may differ depending on whether the test is in the educational/achievement context or if it is more of an attitude/behavioral survey.

In the educational/achievement context, content validation seeks to ensure the items on an exam are appropriate for the content domain being assessed.

A **table of specifications** is a two-way chart which indicates the instructionally relevant learning tasks to be measured. Percentages in the table indicate the relative degree of emphasis that each content area

Let's imagine that I was creating a table of specifications for items on a quiz for this very chapter. The columns represent the types of outcomes we might expect. The American Psychological Association often talks about *KSAs* (knowledge, skills, attitudes), so I will utilize those as a framework. You'll notice that the number of items and percentages do not align mathematically. Because, in the exam, I would likely weight application items (e.g., "work the problem") more highly than knowledge items (e.g., multiple choice, true/false), the relative weighting may differ.

**Table of Specifications**

Learning Objectives	Knowledge	Skills	Attitudes	% of test
Distinguish between different types of validity based on short descriptions.	6 items			30%
Compute and interpret validity coefficients.		2 items		15%
Evaluate the incremental validity of an instrument-of-interest.		1 item		20%
Define and interpret the standard error of estimate.	1 item			15%
Develop a rationale that defends importance of establishing the validity of a measuring instrument.			1 item	20%
<b>TOTALS</b>	7 items	3 items	1 item	100%

**Subject matter experts** (SMEs) are individuals chosen to evaluate items based on their degree of knowledge of the subject being assessed. If SMES are used, the researcher should:

- Report how many SMEs and list their professional qualifications.
- Report any directions the SMEs were given; if they were used to evaluate items, report the extent of agreement.

Empirical procedures for enhancing content validity of educational assessments may include:

- Comparing item-level and total scores with grades; lower grades should get lower scores.
- Analyzing individual errors.
- Observing student work methods (have the students “think aloud” in front of an examiner).
- Evaluating the role of speed, noting how many do not complete the test in the time allowed.
- Correlating the scores with a reading comprehension test (if the exam is highly correlated, then it may be a test of reading and not another subject). Alternatively, if it is a reading comprehension test, give the student the questions (without the passage) to see how well they answered the questions on the basis of prior knowledge.

For surveys and tests outside of educational settings, content validation procedures ask, “Does the test cover a representative sample of the specified skills and knowledge?” and “Is test performance reasonably free from the influence of irrelevant variables?” Naturally, SMEs might be used.

An example of content validation from Industrial-Organizational Psychology is the job analysis which precedes the development of test for employee selection and classification. Not all tests require content analysis. In aptitude and personality tests we are probably more interested in other types of validity evaluation.

#### 4.4.2 Face Validity: The “Un”validity

Face validity is concerned with the question, “How does an assessment look on the ‘face of it’?” Let’s imagine that on a qualification exam for electricians, a math item asks the electrician to estimate the amount of yarn needed to complete a project. The item may be more *face valid* if the calculation was with wire. Thus, face validity can often be improved by reformulating test items in terms that appear relevant and plausible for the context.

Face validity should never be regarded as a substitute for objectively determined validity. In contrast, it should not be assumed that when a (valid and reliable) test has been modified to increase its face validity, that its objective validity and reliability is unaltered. That is, it should be reevaluated.

#### 4.4.3 Criterion-Related Validity

Criterion-related validity has to do with the test’s ability to *predict* an outcome (the criterion). If the criterion is something that occurs simultaneously, it is an assessment of **concurrent validity**; if it is in the future, it is an assessment of **predictive validity**.

A **criterion** is the “thing” that the test should, theoretically, be able to *predict*. That prediction could be occurring at the same time (*concurrent validity*) or at a future time (*predictive validity*). Regardless, the estimate of the criteria must be independent of the survey/assessment being evaluated. The table below provides examples of types of tests and concurrent and predictive validity criteria.

Type of Test	Concurrent Criteria Example	Predictive Criteria Example
A shorter (or cheaper) standardized achievement test	school grades, existing standardized tests	subsequent graduation/college admissions, cumulative GPA
Employee selection tests	decision made by a search committee	subsequent retention or promotion of the selected employee
Assessment of depression severity (shorter or cheaper)	diagnosis from a mental health professional; correlation with an established measure	inpatient hospitalization or act of self-harm

**Contrasted groups** is a specific type of criterion-related validity. Clearly differentiated groups (e.g., sales clerks versus executives; engineers versus musicians) are chosen to see if exam performance or profiles differ in predictable ways.

**Criterion contamination** occurs when test scores, themselves, are used to make decisions about the criteria. To prevent this:

- No person who participates in the assignment of criterion ratings can have any knowledge of the examinee's test scores.
- The test scores must be kept strictly confidential.

There are a number of issues related to criterion-related validity.

- Is the criterion choice appropriate?
  - Criterion validity is only as good as the validity of the criterion to which one is making a comparison.
  - In the 1980s and 1990s there was more attention in this area; that is critics questioned the quality of the criterion being used.
- To what degree can the results of criterion-related validity be generalized?
  - Most tests are developed (intentionally) for a local context, setting, or audience. Consequently, in the local context, the criterion-prediction sample is usually too small (i.e., 50 cases or less).
  - Those who want to generalize the test to a broader population should evaluate the test in relation to the new purpose.
- Is there a role for meta-analysis?
  - Repeated validation studies of our tests, on different samples, results in a number of small-scale studies, each with their own validity coefficients.
  - We can use meta-analytic procedures in reporting the results of validity coefficients when they are used for establishing criterion validity.

#### 4.4.4 Construct Validity

**Construct validity** was introduced in 1954 in the first edition of APA's testing standards and is defined as the extent to which the test may be said to measure a theoretical construct or trait. The overarching focus is on the role of *psychological theory* in test construction and the ability to formulate hypotheses that can be supported (or not) in the evaluation process. Construct validity is established by the accumulation of information from a variety of sources.

There are a number of sources that can be used to support construct validity.

#### 4.4.5 Internal Consistency

In the next **chapter**, you will learn that **internal consistency** is generally considered to be an index of reliability. In the context of criterion-related validity, a goal is to ensure that the criterion is the total score on the test itself. To that end, some of the following could also support this aspect of validity:

- Comparing high and low scorers. Items that fail to show a significantly greater proportion of “passes” in the upper than the lower group are considered invalid, and are modified or eliminated.
- Computing a biserial correlation between the item and total score.
- Correlating the subtest score with the total score. Any subtest whose correlation with the total score is too low is eliminated.

Although some take issue with this notion, the degree of *homogeneity* (the degree to which items assess the same thing) has some bearing on construct validity. There is a tradeoff between items that measure a narrow slice of the construct definition (internal consistency estimates are likely to be higher) and those that sample the construct definition more broadly (internal consistency estimates are likely to be lower).

Admittedly, the contribution of internal consistency data is limited. In absence of external data, it tells us little about **WHAT** the test measures.

#### 4.4.6 Structural Validity

##### 4.4.6.1 Exploratory Factor Analysis

**Exploratory factor analysis** (EFA) is used to simplify the description of behavior by reducing the number of categories (factors or dimensions) to fewer than the number of items. In our research vignette the 6-item Perceptions of Campus Climate Scale will be represented by two factors [Szymanski and Bissonette, 2020] In instrument development, techniques like *principal components analysis* or *principal axis factoring* are used to identify clusters (latent factors) among items. We frequently treat these as scales and subscales.

Imagine the use of 20 tests to 300 people. There would be 190 correlations.

- Irrespective of content, we can probably summarize the intercorrelations of tests with 5-6 factors.

- When the clustering of tests includes vocabulary, analogies, opposites, and sentence completions, we might suggest a “verbal comprehension factor.”
- Factorial validity is the correlation of the test with whatever is common to a group of tests or other indices of behavior. If our single test has a correlation of .66 with the factor on which it loads, then the “factorial validity of the new test as a measure of the common trait is .66.”

When EFA is utilized, the items are “fed” into an iterative process that analyzes the relations and “reveals” (or suggests – we are the ones who interpret the data) how many factors (think scales/subscales) and which items comprise them.

#### 4.4.6.2 Confirmatory Factor Analysis

**Confirmatory factor analysis** (CFA) involves specifying, *a priori*, a proposed relationship of items, scales, and subscales and then testing its *goodness of fit*. In CFA (a form of structural equation modeling [SEM]), the latent variables (usually the higher order scales and total scale score) are positioned to *cause* the responses on the indicators/items.

Subsequent lessons provide examples of both EFA and CFA approaches to psychometrics.

#### 4.4.7 Experimental Interventions

Construct validity is also supported by hypothesis testing and experimentation. If we expect that the construct assessed by the instrument is malleable (e.g., depression) and that an intervention could change it, then a random clinical trial that evaluated the effectiveness of an intervention (and it worked – depression scores declined) would simultaneously provide support for the intervention as well as the instrument.

#### 4.4.8 Convergent and Discriminant Validity

In a psychometric evaluation, we will often administer our instrument-of-interest along with a battery of instruments that are more-and-less related. **Convergent validity** is supported when there are *moderately high* correlations between our tests and the instruments with which we expect moderately high correlations. In contrast, **discriminant validity** is established by low and/or non-significant correlations between our instrument-of-interest and instruments that should be unrelated. For example, we want a low and non-significant correlation between a quantitative reasoning test and scores on a reading comprehension test. Why? Because if the correlation is too high, the test cannot discriminate between reading comprehension and math.

There are no strict cut-offs to establish convergence or discrimination. We can even ask, “Could a correlation intended to support convergence be too high?” It is possible! Unless the instrument-of-interest offers advantages such as brevity or cost, then correlations that fall into the ranges of multicollinearity or singularity can indicate unnecessary duplication or redundancy.

In our research vignette, Szymanski and Bissonette [2020] conducted a correlation matrix that reports the bivariate relations between the LGBTQ Campus Climate full-scale as well as the College Response and Stigma subscales with measures that assess (a) LGBTQ victimization, (b) satisfaction with college, (c) persistence attitudes, and (d) anxiety, and (e) depression.

In order to produce this correlation matrix, we must first score each of the scales. In the items we prepared, the Perceptions of LGBTQ Campus Climate scale had one reverse-scored item. Similarly, the Institutional Goals and Commitments Scale had three reversed items. A first step in scoring is reversing these items.

The naming conventions that researchers use vary. I added an “NR” (for “needs reversing) to the original items so that I would remember to reverse-score them. I also am careful to reverse-score items into new variables. Otherwise, we risk getting confused about whether/not items are in their original or reversed formats.

Reverse-scoring the item is easily accomplished by subtracting the variable from “one plus” the scaling. Because both of these scales were on a 7-point scale, we will subtract the “NR” variablse from 8.

```
# Reverse scoring the single item from the LGBTQ Campus Climate Scale
dfSzy <- dfSzy %>%
  dplyr::mutate(unsupported = 8 - supportiveNR)

# Reversing three items on the Institutional and Goals Commitments
# scale

dfSzy <- dfSzy %>%
  dplyr::mutate(not_graduate = 8 - not_graduateNR) %>%
  dplyr::mutate(undecided = 8 - undecidedNR) %>%
  dplyr::mutate(grades_unimportant = 8 - grades_unimportantNR)
```

Next we create scale and/or subscale scores. The *sjstats::mean\_n()* function allows us to specify how many items (whole number) or what percentage of items should be present in order to get the mean. It is customary to require 75-80% of items to be present for scoring. Three-item variables might allow one missing (i.e., 66%). In the code below, I first make lists of the variables that belong in each scale and subscale. then I create th enew variables.

```
# Making the list of variables
LGBTQ_Climate <- c("cold", "unresponsive", "unsupported", "negative",
  "heterosexism", "harassed")
CollResponse <- c("cold", "unresponsive", "unsupported")
Stigma <- c("negative", "heterosexism", "harassed")
Victimization <- c("Vic1", "Vic2", "Vic3", "Vic4", "Vic5", "Vic6", "Vic7",
  "Vic8", "Vic9")
CampSat <- c("Sat1", "Sat2", "Sat3", "Sat4", "Sat5")
Persist <- c("graduation_importance", "right_decision", "will_register",
  "not_graduate", "undecided", "grades_unimportant")
GAD7 <- c("nervous", "worry_control", "much_worry", "cant_relax", "restless",
  "irritable", "afraid")
PHQ9 <- c("anhedonia", "down", "sleep", "lo_energy", "appetite", "selfworth",
  "concentration", "too_slowfast", "s_ideation")

# Creating the new variables
dfSzy$LGBTQclimate <- sjstats::mean_n(dfSzy[, LGBTQ_Climate], 0.75)
```

```

dfSzy$CollegeRx <- sjstats::mean_n(dfSzy[, CollResponse], 0.66)
dfSzy$Stigma <- sjstats::mean_n(dfSzy[, Stigma], 0.66)
dfSzy$Victimization <- sjstats::mean_n(dfSzy[, Victimization], 0.8)
dfSzy$CampusSat <- sjstats::mean_n(dfSzy[, CampSat], 0.75)
dfSzy$Persistence <- sjstats::mean_n(dfSzy[, Persist], 0.8)
dfSzy$Anxiety <- sjstats::mean_n(dfSzy[, GAD7], 0.75)
dfSzy$Depression <- sjstats::mean_n(dfSzy[, PHQ9], 0.8)

# If the scoring code above does not work for you, try the format
# below which involves inserting two periods in front of the variable
# list. One example is provided. dfLewis$Belonging <-
# sjstats::mean_n(dfLewis[, ..Belonging_vars], 0.80)

```

A correlation matrix of the scaled scores allows us to compare the our scale(s) of interest to others within its nomological net.

```

apaTables::apa.cor.table(dfSzy[c("LGBTQclimate", "CollegeRx", "Stigma",
                                "Victimization", "CampusSat", "Persistence", "Anxiety", "Depression")],
                          filename = "SzyCor.doc", table.number = 1, show.sig.stars = TRUE, landscape = TRUE)

```

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. LGBTQclimate	4.00	0.63			
2. CollegeRx	4.04	0.77	.83** [.81, .85]		
3. Stigma	3.96	0.76	.83** [.80, .85]	.37** [.31, .44]	
4. Victimization	1.55	0.33	.01 [-.06, .09]	-.17** [-.25, -.10]	.20** [.13, .27]
5. CampusSat	4.24	0.70	-.49** [-.55, -.43]	-.46** [-.52, -.40]	-.35** [-.41, -.28]
6. Persistence	3.03	0.42	-.21** [-.28, -.13]	-.17** [-.25, -.10]	-.17** [-.25, -.10]
7. Anxiety	1.49	0.38	.17** [.10, .25]	.12** [.04, .19]	.17** [.09, .24]

8. Depression	1.52	0.29	.18**	.14**	.15**
			[.10, .25]	[.07, .22]	[.08, .23]

4	5	6	7
---	---	---	---

-.17\*\*  
[-.25, -.10]

-.04            .34\*\*  
[-.11, .04]    [.27, .40]

.15\*\*           -.20\*\*           -.10\*  
[.07, .23]      [-.28, -.13]    [-.18, -.02]

.15\*\*           -.23\*\*           -.10\*\*           .54\*\*  
[.08, .23]      [-.30, -.15]    [-.18, -.03]    [.48, .59]

Note. M and SD are used to represent mean and standard deviation, respectively.  
Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations  
that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

Examination of these values follow some expected patterns. First, the LGBTQ climate score (i.e., the total scale score) is highly correlated with each of its subscales (College Response  $r = .83$ ,  $p < 0.01$ ; Stigma  $r = .83$ ,  $p = 0.01$ ). These strong correlations are somewhat misleading because half of the items on the total scale are the same items on each of the subscales. The correlation between the two subscales is still statistically significant, but much lower ( $r = 0.37$ ,  $p = 0.01$ ).

Convergent and discriminant validity are of interest when we compare the LGBTQ Climate total scale score and the College Response and Stigma subscales with the additional measures. Regarding the total LGBTQ Climate score, a very strong correlation was observed with campus satisfaction ( $r = -0.49$ ,  $p < 0.01$ ); less strong correlations were observed with and persistence ( $r = -0.21$ ,  $p < 0.01$ ), anxiety ( $r = 0.17$ ,  $p < 0.01$ ), and depression ( $r = 0.18$ ,  $p < 0.01$ ). Recalling that higher scores on the LGBTQ Campus Climate score indicate a negative climate, we see that as the

LGBTQ campus climate becomes increasingly stigmatizing and nonresponsive, students experience lower overall campus satisfaction and are less likely to persist at that institution. The correlation between LGBTQ Campus Climate and victimization was non-significant ( $r = 0.01, p > 0.05$ ).

In assessing patterns of convergent and discriminant validity, the researcher would also take the time to map out the subscales (i.e., College Response, Stigma) with the additional measures.

#### 4.4.8.1 Determining Statistically Significant Differences Between Correlations

Without a formal test, it is inappropriate for researchers to declare that one correlation is stronger than another. The package *cocor* allows the comparisons of *dependent* (i.e., all respondents are from the same sample) and *independent* (i.e., correlations are compared between two different samples) where the correlations themselves can be *overlapping* (i.e., with one common variable) or *non-overlapping* (i.e., the variables in both correlations are different).

Because all of the correlations we computed are within the same sample, they are *dependent*. When assessing convergent and discriminant validity it is common to ask if the correlations between the additional measures are different between the subscales of the focal measure. Results could support the notion that the subscales are related and yet different. An example of this might be comparing the correlations between victimization with the College Response subscale ( $r = -.17, p < 0.01$ ) and victimization with Stigma subscale ( $r = 0.20, p < 0.01$ ). This test would be *overlapping* because the victimization variable is in common. Here's the code:

```
cocor::cocor(formula = ~CollegeRx + Victimization | Stigma + Victimization,
             data = dfSzy)
```

Results of a comparison of two overlapping correlations based on dependent groups

```
Comparison between r.jk (Victimization, CollegeRx) = -0.1741 and r.jh (Victimization, Stigma) =
Difference: r.jk - r.jh = -0.3756
Related correlation: r.kh = 0.3734
Data: dfSzy: j = Victimization, k = CollegeRx, h = Stigma
Group size: n = 646
Null hypothesis: r.jk is equal to r.jh
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)
Alpha: 0.05
```

```
pearson1898: Pearson and Filon's z (1898)
z = -8.9455, p-value = 0.0000
Null hypothesis rejected
```

```
hotelling1940: Hotelling's t (1940)
t = -9.0342, df = 643, p-value = 0.0000
Null hypothesis rejected
```

```
williams1959: Williams' t (1959)
t = -9.0340, df = 643, p-value = 0.0000
```

Null hypothesis rejected

olkin1967: Olkin's z (1967)

$z = -8.9455$ , p-value = 0.0000  
Null hypothesis rejected

dunn1969: Dunn and Clark's z (1969)

$z = -8.7137$ , p-value = 0.0000  
Null hypothesis rejected

hendrickson1970: Hendrickson, Stanley, and Hills' (1970) modification of Williams' t (1959)

$t = -9.0341$ , df = 643, p-value = 0.0000  
Null hypothesis rejected

steiger1980: Steiger's (1980) modification of Dunn and Clark's z (1969) using average correlation

$z = -8.6124$ , p-value = 0.0000  
Null hypothesis rejected

meng1992: Meng, Rosenthal, and Rubin's z (1992)

$z = -8.5080$ , p-value = 0.0000  
Null hypothesis rejected  
95% confidence interval for  $r_{jk} - r_{jh}$ : -0.4678 -0.2926  
Null hypothesis rejected (Interval does not include 0)

hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) using

$z = -8.6124$ , p-value = 0.0000  
Null hypothesis rejected

zou2007: Zou's (2007) confidence interval

95% confidence interval for  $r_{jk} - r_{jh}$ : -0.4568 -0.2921  
Null hypothesis rejected (Interval does not include 0)

Fisher's z-test ( $z = -8.6124$ ,  $p < 0.001$ ) tells us that the correlations are statistically significantly different from each other; Zhou's confidence intervals provide the CI for the size of the difference between the two correlations. That is, the difference could be as large as -0.4568 or as small as -0.2921.

Another type of correlation comparison is with a the total and/or subscales, looking at the relative magnitude of their correlation with different variables. For example, we might wish to ask if the LGBTQ Campus Climate total scale is different degrees of correlation with anxiety ( $r = 0.17$ ,  $p < 0.01$ ) and depression ( $r = .18$ ,  $p < 0.01$ ).

```
cocor::cocor(formula = ~LGBTQclimate + Anxiety | LGBTQclimate + Depression,
  data = dfSzy)
```

Results of a comparison of two overlapping correlations based on dependent groups

Comparison between r.jk (LGBTQclimate, Anxiety) = 0.1724 and r.jh (LGBTQclimate, Depression) = Difference: r.jk - r.jh = -0.0071  
Related correlation: r.kh = 0.5387  
Data: dfSzy: j = LGBTQclimate, k = Anxiety, h = Depression  
Group size: n = 646  
Null hypothesis: r.jk is equal to r.jh  
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)  
Alpha: 0.05

pearson1898: Pearson and Filon's z (1898)  
z = -0.1914, p-value = 0.8482  
Null hypothesis retained

hotelling1940: Hotelling's t (1940)  
t = -0.1911, df = 643, p-value = 0.8485  
Null hypothesis retained

williams1959: Williams' t (1959)  
t = -0.1909, df = 643, p-value = 0.8487  
Null hypothesis retained

olkin1967: Olkin's z (1967)  
z = -0.1914, p-value = 0.8482  
Null hypothesis retained

dunn1969: Dunn and Clark's z (1969)  
z = -0.1909, p-value = 0.8486  
Null hypothesis retained

hendrickson1970: Hendrickson, Stanley, and Hills' (1970) modification of Williams' t (1959)  
t = -0.1911, df = 643, p-value = 0.8485  
Null hypothesis retained

steiger1980: Steiger's (1980) modification of Dunn and Clark's z (1969) using average correlation  
z = -0.1909, p-value = 0.8486  
Null hypothesis retained

meng1992: Meng, Rosenthal, and Rubin's z (1992)  
z = -0.1909, p-value = 0.8486  
Null hypothesis retained  
95% confidence interval for r.jk - r.jh: -0.0825 0.0678  
Null hypothesis retained (Interval includes 0)

hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) using  
z = -0.1909, p-value = 0.8486  
Null hypothesis retained

zou2007: Zou's (2007) confidence interval

```
95% confidence interval for r.jk - r.jh: -0.0798 0.0656
Null hypothesis retained (Interval includes 0)
```

Fisher's z-test ( $z = -0.1909$ ,  $p = 0.8486$ ) tells us that the correlations are not statistically significantly different from each other; Zhou's confidence intervals indicate that the differences range between -0.0798 and 0.0656. Because this interval crosses zero, we know that the difference could be zero, or reversed in direction.

#### 4.4.8.2 Multitrait-Multimethod Matrix

The **multitrait-multimethod matrix** is a systematic experimental design for the dual approach of convergent and discriminant validation, which requires the assessment of two or more traits (classically, math, English, and reading scores) by two more methods (self, parent, and teacher). Conducting a web-based image search on this term will show a matrix of alpha coefficients and correlation coefficients that are interpreted in relationship to each other. Roughly:

- alpha coefficients (internal consistency) should be the highest,
- validity coefficients (correlations of the same trait assessed by different methods) should be higher than correlations between different traits measured by different methods,
- validity coefficients (correlations of the same trait assessed by different methods) should be higher than different traits measured by the same method.

#### 4.4.9 Incremental Validity

**Incremental validity** is the increase in predictive validity attributable to the test. It indicates the contribution the test makes to the selection of individuals who will meet the minimum standards in criterion performance. There are different ways to assess this – one of the most common is to first enter known predictors and then see if the instrument-of-interest continues to account variance over-and-above those that are entered.

In the Szymanski and Bissonette [2020] psychometric evaluation, the negative relations with satisfaction with college and intention to persist in college as well as positive relations with both anxiety and depression persisted even after controlling for LGBTQ victimization experiences.

I will demonstrate this procedure, predicting the contribution that the LGBTQ Campus Climate total scale score has on predicting intention to persist in college, over and above LGBTQ victimization.

The process is to use hierarchical linear regression. Two models are built. In the first mode (“PfV” stands [in my mind] for “Persistence from Victimization”), persistence is predicted from victimization. The second model adds the LGBTQ Campus Climate Scale. I asked for summaries of each model. Then the *anova()* function compares the model.

```
PfV <- lm(Persistence ~ Victimization, data = dfSzy)
PfVC <- lm(Persistence ~ Victimization + LGBTQclimate, data = dfSzy)
summary(PfV)
```

Call:

```
lm(formula = Persistence ~ Victimization, data = dfSzy)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.19281	-0.34150	-0.02281	0.29669	1.32226

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.09906	0.07947	38.997 <0.0000000000000002 ***	
Victimization	-0.04566	0.05023	-0.909	0.364
---				
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1	' '	1	

Residual standard error: 0.4224 on 644 degrees of freedom

Multiple R-squared: 0.001281, Adjusted R-squared: -0.0002694

F-statistic: 0.8263 on 1 and 644 DF, p-value: 0.3637

From the PfV model we learn that victimization has a non-significant effect on intentions to persist in college ( $B = -0.046$ ,  $p = 0.364$ ). Further, the  $R^2$  is quite small (0.001).

```
summary(PfVC)
```

Call:

```
lm(formula = Persistence ~ Victimization + LGBTQclimate, data = dfSzy)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.1696	-0.2842	0.0094	0.2569	1.3571

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.64440	0.12795	28.483 < 0.0000000000000002 ***	
Victimization	-0.04183	0.04919	-0.850	0.395
LGBTQclimate	-0.13788	0.02568	-5.369	0.000000111 ***
---				
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1	' '	1	

Residual standard error: 0.4135 on 643 degrees of freedom

Multiple R-squared: 0.04413, Adjusted R-squared: 0.04116

F-statistic: 14.84 on 2 and 643 DF, p-value: 0.0000004991

In the PfVC model, we see that the LGBTQ Campus Climate full scale score has a significant impact on intentions to persist. Specifically, for each additional point higher on the LGBTQ climate score, intentions to persist decrease by .14 points ( $p < 0.001$ ). Together, the model accounts for 4% of the variance, representing a  $\Delta R^2$  of 4%.

```
# calculating R2 change
0.04413 - 0.001281
```

```
[1] 0.042849
```

```
anova(PfV, PfVC)
```

Analysis of Variance Table

```
Model 1: Persistence ~ Victimization
Model 2: Persistence ~ Victimization + LGBTQclimate
Res.Df   RSS Df Sum of Sq    F    Pr(>F)
1     644 114.88
2     643 109.95  1      4.929 28.824 0.0000001108 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that there is a statistically significant difference between the models  $F(1, 643) = 28.824, p < 0.001$ .

Let's try another model. With anxiety as our dependent variable, the code below asks if LGBTQ Campus Climate accounts for a proportion of the variance over-and-above victimization.

```
AfV <- lm(Anxiety ~ Victimization, data = dfSzy)
AfVC <- lm(Anxiety ~ Victimization + LGBTQclimate, data = dfSzy)
summary(AfV)
```

Call:

```
lm(formula = Anxiety ~ Victimization, data = dfSzy)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.05943	-0.30935	-0.01528	0.31306	1.24148

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.21756	0.07131	17.073	< 0.0000000000000002 ***
Victimization	0.17427	0.04508	3.866	0.000122 ***

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.379 on 644 degrees of freedom

Multiple R-squared: 0.02268, Adjusted R-squared: 0.02116

F-statistic: 14.95 on 1 and 644 DF, p-value: 0.0001219

```
summary(AfVC)
```

Call:

```
lm(formula = Anxiety ~ Victimization + LGBTQclimate, data = dfSzy)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.00814	-0.29249	-0.02563	0.29877	1.20705

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.81071	0.11561	7.012	0.00000000000595 ***
Victimization	0.17141	0.04444	3.857	0.000126 ***
LGBTQclimate	0.10287	0.02320	4.433	0.00001092659570 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3736 on 643 degrees of freedom

Multiple R-squared: 0.05166, Adjusted R-squared: 0.04872

F-statistic: 17.52 on 2 and 643 DF, p-value: 0.0000000392

```
anova(AfV, AfVC)
```

Analysis of Variance Table

Model 1: Anxiety ~ Victimization

Model 2: Anxiety ~ Victimization + LGBTQclimate

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	644	92.513			
2	643	89.770	1	2.7436	19.651 0.00001093 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This model is a little more exciting in that our first model (AfV) is statistically significant ( $B = 0.174, p < 0.001$ ). That is, victimization has a statistically significant effect on anxiety, accounting for 2% of the variance. In the second model, the LGBTQ Campus Climate total scale score is also significant ( $B = 0.103, p < 0.001$ ), and accounts for an additional 3% of variance ( $\Delta R^2 = 0.029$ ). There is a statistically significant difference between models ( $F[1, 643] = 19.651, p < .001$ ).

```
# calculating change in R2
0.05166 - 0.02268
```

[1] 0.02898

#### 4.4.10 Considering the Individual and Social Consequences of Testing

Messick [Messick, 2000] and others recommend that the “consequences of testing” be included in the concept of test validity. Messick’s point was to consider the unintended consequences of specific uses. That is, their use may be detrimental to individuals or to members of certain ethnic or other populations with diverse experiential backgrounds. Examples of inappropriate use have included:

- The California Psychological Inventory (CPI) being used as a screening tool for employment as a security job. Two of its items inquired about same-sex activities and the employer was using this to screen out gay men. Applicants were able to demonstrate, in court, a consistent rejection of gay applicants.
- While this is not a psychological test, urine samples are often collected as drug screening tools. In reality, urine can reveal a number of things, such as pregnancy.

The issue begs “conflicting goals.” In this case, the problem was not caused by the test but rather by its misuse. Studying the “consequences” of testing is one that is not necessarily answerable by empirical data/statistical analysis. It requires critical observation, human judgment, and systematic debate.

## 4.5 Factors Affecting Validity Coefficients

Keeping in mind that a *validity coefficient* is merely the correlation between the test and some criteria, the same elements that impact the magnitude and significance of a correlation coefficient will similarly effect a validity coefficient.

**Nature of the group.** A test that has high validity in predicting a particular criterion in one population, may have little or no validity in predicting the same criterion in another population. If a test is designed for use in diverse populations, information about the population generalizability should be reported in the technical manuals.

**Sample heterogeneity.** Other things being equal, if there is a linear relationship between X and Y, it will have a greater magnitude when the sample is heterogeneous.

**Pre-selection.** Just like internal and external validity in a research design can be threatened by selection issues, pre-selection can also impact the validity coefficients of a measure. For example, if we are evaluating a new test for job selection, we may select a group of newly hired employees. We plan to collect some measure of job performance at a later date. Our results may be limited by the criteria used to select the employees. Were they volunteers? Were they only those hired? Were they ALL of the applicants?

**Validity coefficients may change over time.** Consider the relationship between the college boards and grade point average at Yale University. Fifty years ago,  $r_{xy} = .72$ ; today  $r_{xy} = .52$ . Why? The nature of the student body has become more diverse (50 years ago, the student body was predominantly White, high SES, and male).

The **form of the relationship** matters. The Pearson R assumes the relationship between the predictor and criterion variables is linear, uniform, and homoschedastic (equal variability throughout

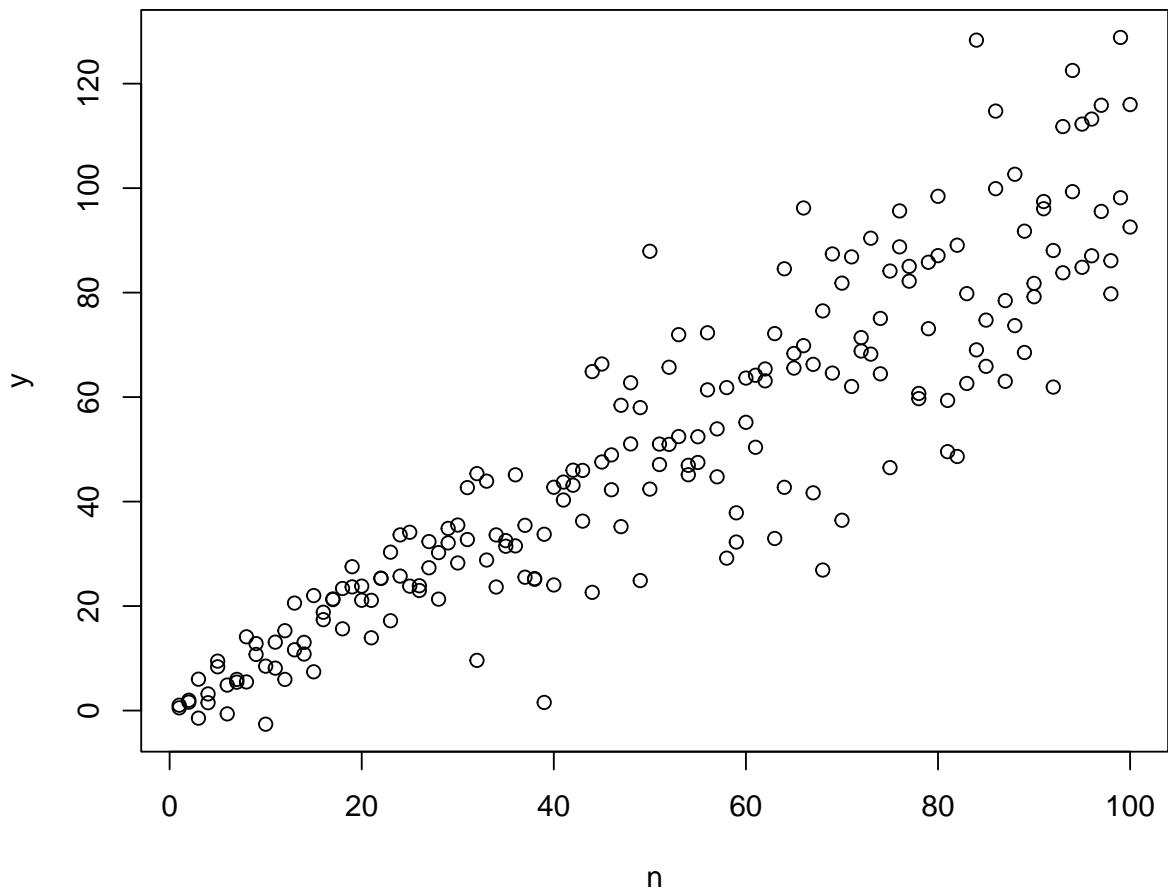


Figure 4.1: Illustration of heteroscedasticity

the range of a bivariate distribution). When the variability is unequal throughout the range of the distribution the relationship is heteroscedastic.

There could also be other factors involved in the relationship between the instrument and the criterion:

- curvilinearity
- an undetected mechanism, such as a moderator

Finally, what is our threshold for acceptability?

- Consider statistical significance – but also its limitations (e.g., power, Type I error, Type II error)
- Consider the magnitude of the correlation; and also  $R^2$  (the proportion of variance accounted for)
- Consider error:
  - The standard error of the estimate shows the margin of error to be expected in the individuals predicted criterion score as the result of the imperfect validity of the instrument.

$$SE_{est} = SD_y \sqrt{1 - r_{xy}^2}$$

Where  $r_{xy}^2$  is the square of the validity coefficient  $SD_y$  is the standard deviation of the criterion scores

If the validity were perfect ( $r_{xy}^2 = 1.00$ ), the error of estimate would be 0.00. If the validity were zero, the error of estimate would equal  $SD_y$ .

Interpreting  $SE_{est}$

If  $r_{xy} = .80$ , then  $\sqrt{1 - r_{xy}^2} = .60$

Error is 60% as large as it would be by chance. Stated another way, predicting an individual's criterion performance has a margin of error that is 40% smaller than it would be by chance.

To obtain the  $SE_{est}$ , we merely multiply by the  $SD_y$ . This puts error in the metric of the criterion variable.

Your Turn If  $r_{xy} = .25$ , then  $\sqrt{1 - r_{xy}^2} = ??$

Make a statement about chance. Make a statement about margin of error.

## 4.6 Practice Problems

In each of these lessons, I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. With each of these options, I encourage you to interpret examine aspects of the construct validity through the creation and interpretation of validity coefficients. Ideally, you will examine both convergent/discriminant validity as well as incremental validity.

#### 4.6.1 Problem #1: Play around with this simulation.

Copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results.

If calculating is new to you, perhaps you just change the number in “set.seed(210907)” from 210907 to something else. Your results should parallel those obtained in the lecture, making it easier for you to check your work as you go.

#### 4.6.2 Problem #2: Conduct the reliability analysis selecting different variables.

The Szymanski and Bissonette [2020] article conducted a handful of incremental validity assessments. Select different outcome variables (e.g., depression) and/or use the subscales as the instrument-of-interest.

#### 4.6.3 Problem #3: Try something entirely new.

Using data for which you have permission and access (e.g., IRB approved data you have collected or from your lab; data you simulate from a published article; data from an open science repository; data from other chapters in this OER), create validity coefficients and use three variables to estimate the incremental validity of the instrument-of-interest.

#### 4.6.4 Grading Rubric

Assignment Component	Points Possible	Points Earned
1. Check and, if needed, format and score the data.	5	_____
2. Create a correlation matrix that includes the instrument-of-interest and the variables that will have varying degrees of relation.	5	_____
3. With convergent and discriminant validity in mind, interpret the validity coefficients; this should include an assessment about whether the correlation coefficients (at least two different pairings) are statistically significantly different from each other.	5	_____
4. With at least three variables, evaluate the degree to which the instrument demonstrates incremental validity (this should involve two regression equations and their statistical comparison).	5	_____
5. Explanation to grader.	5	_____
<b>Totals</b>	<b>25</b>	_____

## 4.7 Homeworked Example

[Screencast Link](#)

For more information about the data used in this homeworked example, please refer to the description and codebook located at the end of the [introduction](#) in first volume of ReCentering Psych Stats.

As a brief review, this data is part of an IRB-approved study, with consent to use in teaching demonstrations and to be made available to the general public via the open science framework. Hence, it is appropriate to use in this context. You will notice there are student- and teacher- IDs. These numbers are not actual student and teacher IDs, rather they were further re-identified so that they could not be connected to actual people.

Because this is an actual dataset, if you wish to work the problem along with me, you will need to download the [ReC.rds](#) data file from the Worked\_Examples folder in the ReC\_Psychometrics project on the GitHub.

The course evaluation items can be divided into three subscales:

- **Valued by the student** includes the items: ValObjectives, IncrUnderstanding, IncrInterest
- **Traditional pedagogy** includes the items: ClearResponsibilities, EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation
- **Socially responsive pedagogy** includes the items: InclusvClassrm, EquitableEval, Multi-Perspectives, DEIIntegration

In this homework focused on validity we will score the total scale and subscales, create a correlation matrix of our scales with a different scale (or item), formally test to see if correlation coefficients are statistically significantly different from each other, conduct a test of incremental validity.

### 4.7.1 Check and, if needed, format data

```
big <- readRDS("ReC.rds")
```

Let's check the structure...

```
str(big)
```

```
Classes 'data.table' and 'data.frame': 310 obs. of 33 variables:
 $ deID           : int 1 2 3 4 5 6 7 8 9 10 ...
 $ CourseID       : int 57085635 57085635 57085635 57085635 57085635 57085635 57085635 ...
 $ Dept           : chr "CPY" "CPY" "CPY" "CPY" ...
 $ Course          : Factor w/ 3 levels "Psychometrics",...: 2 2 2 2 2 2 2 2 2 2 ...
 $ StatsPkg        : Factor w/ 2 levels "SPSS","R": 2 2 2 2 2 2 2 2 2 2 ...
 $ Centering       : Factor w/ 2 levels "Pre","Re": 2 2 2 2 2 2 2 2 2 2 ...
 $ Year            : int 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ Quarter         : chr "Fall" "Fall" "Fall" "Fall" ...
```

```

$ IncrInterest          : int  5 3 4 2 4 3 5 3 2 5 ...
$ IncrUnderstanding     : int  2 3 4 3 4 4 5 2 4 5 ...
$ ValObjectives          : int  5 5 4 4 5 5 5 5 4 5 ...
$ ApprAssignments        : int  5 4 4 4 5 3 5 3 3 5 ...
$ EffectiveAnswers       : int  5 3 5 3 5 3 4 3 2 3 ...
$ Respectful             : int  5 5 4 5 5 4 5 4 5 5 ...
$ ClearResponsibilities : int  5 5 4 4 5 4 5 4 4 5 ...
$ Feedback                : int  5 3 4 2 5 NA 5 4 4 5 ...
$ OvInstructor            : int  5 4 4 3 5 3 5 4 3 5 ...
$ MultPerspectives         : int  5 5 4 5 5 4 5 5 5 5 ...
$ OvCourse                : int  3 4 4 3 5 3 5 3 2 5 ...
$ InclusvClassrm          : int  5 5 5 5 5 4 5 5 4 5 ...
$ DEIintegration           : int  5 5 5 5 5 4 5 5 5 5 ...
$ ClearPresentation        : int  4 4 4 2 5 3 4 4 4 5 ...
$ ApprWorkload             : int  5 5 3 4 4 2 5 4 4 5 ...
$ MyContribution           : int  4 4 4 4 5 4 4 3 4 5 ...
$ InspiredInterest         : int  5 3 4 3 5 3 5 4 4 5 ...
$ Faith                   : int  5 NA 4 2 NA NA 4 4 4 NA ...
$ EquitableEval            : int  5 5 3 5 5 3 5 5 3 5 ...
$ SPFC.Decolonize.Opt.Out: chr  "" "" "" ...
$ ProgramYear              : Factor w/ 3 levels "Second","Transition",...: 3 3 3 3 3 3 3 3 3 ...
$ ClearOrganization         : int  3 4 3 4 4 4 5 4 4 5 ...
$ RegPrepare               : int  5 4 4 4 4 3 4 4 4 5 ...
$ EffectiveLearning          : int  2 4 3 4 4 2 5 3 2 5 ...
$ AccessibleInstructor      : int  5 4 4 4 5 4 5 4 5 5 ...
- attr(*, ".internal.selfref")=<externalptr>

```

We will need to create the three subscales. The codebook above, lists which variables go in each subscale score.

```

# Making the list of variables
ValuedVars <- c("ValObjectives", "IncrUnderstanding", "IncrInterest")
TradPedVars <- c("ClearResponsibilities", "EffectiveAnswers", "Feedback",
                 "ClearOrganization", "ClearPresentation")
SRPedVars <- c("InclusvClassrm", "EquitableEval", "MultPerspectives", "DEIintegration")
Total <- c("ValObjectives", "IncrUnderstanding", "IncrInterest", "ClearResponsibilities",
          "EffectiveAnswers", "Feedback", "ClearOrganization", "ClearPresentation",
          "InclusvClassrm", "EquitableEval", "MultPerspectives", "DEIintegration")

# Creating the new variables
big$Valued <- sjstats::mean_n(big[, ValuedVars], 0.66)
big$TradPed <- sjstats::mean_n(big[, TradPedVars], 0.75)
big$SRPed <- sjstats::mean_n(big[, SRPedVars], 0.75)
big$Total <- sjstats::mean_n(big[, Total], 0.8)

# If the scoring code above does not work for you, try the format
# below which involves inserting two periods in front of the variable

```

```
# list. One example is provided. dfLewis$Belonging <-
# sjstats::mean_n(dfLewis[, ..Belonging_vars], 0.80)
```

#### 4.7.2 Create a correlation matrix that includes the instrument-of-interest and the variables that will have varying degrees of relation

Unfortunately, data from the course evals don't include any outside scales. However, I didn't include the "Overall Instructor" (OvInstructor) in any of the items, so we *could* think of it as a way to look at convergent and discriminant validity.

```
apaTables::apa.cor.table(big[c("Valued", "TradPed", "SRPed", "OvInstructor")],
  filename = "ReC_cortable.doc", table.number = 1, show.sig.stars = TRUE,
  landscape = TRUE)
```

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. Valued	4.25	0.68			
2. TradPed	4.25	0.76	.70** [.63, .75]		
3. SRPed	4.52	0.58	.56** [.48, .64]	.71** [.65, .76]	
4. OvInstructor	4.37	0.94	.63** [.56, .70]	.80** [.76, .84]	.67** [.60, .73]

Note. M and SD are used to represent mean and standard deviation, respectively.  
 Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

All the correlations are strong and positive, but look at the correlation between Overall Instructor and SCRPed!

#### 4.7.3 With convergent and discriminant validity in mind, interpret the validity coefficients; this should include an assessment about whether the correlation coefficients (at least two different pairings) are statistically significantly different from each other.

We need to see if these correlations are statistically significantly different from each other. I am interested in knowing if the correlations between Overall Instructor and each of the three course dimensions (Valued [ $r = 0.63, p < 0.01$ ], TradPed [ $r = 0.80, p < 0.01$ ], SRPed [ $r = 0.67, p < 0.01$ ]) are statistically significantly different from each other.

```
cocor::cocor(formula = ~Valued + OvInstructor | TradPed + OvInstructor,
  data = big)
```

Results of a comparison of two overlapping correlations based on dependent groups

```
Comparison between r.jk (OvInstructor, Valued) = 0.6344 and r.jh (OvInstructor, TradPed) = 0.7952
Difference: r.jk - r.jh = -0.1652
Related correlation: r.kh = 0.697
Data: big: j = OvInstructor, k = Valued, h = TradPed
Group size: n = 307
Null hypothesis: r.jk is equal to r.jh
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)
Alpha: 0.05
```

```
pearson1898: Pearson and Filon's z (1898)
z = -5.4939, p-value = 0.0000
Null hypothesis rejected
```

```
hotelling1940: Hotelling's t (1940)
t = -6.2651, df = 304, p-value = 0.0000
Null hypothesis rejected
```

```
williams1959: Williams' t (1959)
t = -6.1447, df = 304, p-value = 0.0000
Null hypothesis rejected
```

```
olkin1967: Olkin's z (1967)
z = -5.4939, p-value = 0.0000
Null hypothesis rejected
```

```
dunn1969: Dunn and Clark's z (1969)
z = -5.9983, p-value = 0.0000
Null hypothesis rejected
```

```
hendrickson1970: Hendrickson, Stanley, and Hills' (1970) modification of Williams' t (1959)
t = -6.2651, df = 304, p-value = 0.0000
```

Null hypothesis rejected

```
steiger1980: Steiger's (1980) modification of Dunn and Clark's z (1969) using average correlation
z = -5.9444, p-value = 0.0000
Null hypothesis rejected
```

```
meng1992: Meng, Rosenthal, and Rubin's z (1992)
z = -5.9182, p-value = 0.0000
Null hypothesis rejected
95% confidence interval for r.jk - r.jh: -0.4644 -0.2333
Null hypothesis rejected (Interval does not include 0)
```

```
hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) using average correlation
z = -5.8868, p-value = 0.0000
Null hypothesis rejected
```

```
zou2007: Zou's (2007) confidence interval
95% confidence interval for r.jk - r.jh: -0.2282 -0.1089
Null hypothesis rejected (Interval does not include 0)
```

Fisher's z-test ( $z = -5.887$ ,  $p < 0.001$ ) indicates that the correlation of overall instructor with the valued subscale ( $r = 0.63$ ) is lower than its correlation with the traditional pedagogy subscale ( $r = 0.80$ ).

```
cocor::cocor(formula = ~TradPed + OvInstructor | SRPed + OvInstructor,
  data = big)
```

Results of a comparison of two overlapping correlations based on dependent groups

```
Comparison between r.jk (OvInstructor, TradPed) = 0.7962 and r.jh (OvInstructor, SRPed) = 0.6785
Difference: r.jk - r.jh = 0.1211
Related correlation: r.kh = 0.7091
Data: big: j = OvInstructor, k = TradPed, h = SRPed
Group size: n = 298
Null hypothesis: r.jk is equal to r.jh
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)
Alpha: 0.05
```

```
pearson1898: Pearson and Filon's z (1898)
z = 4.2785, p-value = 0.0000
Null hypothesis rejected
```

```
hotelling1940: Hotelling's t (1940)
t = 4.6684, df = 295, p-value = 0.0000
Null hypothesis rejected
```

williams1959: Williams' t (1959)  
 t = 4.5800, df = 295, p-value = 0.0000  
 Null hypothesis rejected

olkin1967: Olkin's z (1967)  
 z = 4.2785, p-value = 0.0000  
 Null hypothesis rejected

dunn1969: Dunn and Clark's z (1969)  
 z = 4.5174, p-value = 0.0000  
 Null hypothesis rejected

hendrickson1970: Hendrickson, Stanley, and Hills' (1970) modification of Williams' t (1959)  
 t = 4.6684, df = 295, p-value = 0.0000  
 Null hypothesis rejected

steiger1980: Steiger's (1980) modification of Dunn and Clark's z (1969) using average correlation  
 z = 4.4945, p-value = 0.0000  
 Null hypothesis rejected

meng1992: Meng, Rosenthal, and Rubin's z (1992)  
 z = 4.4834, p-value = 0.0000  
 Null hypothesis rejected  
 95% confidence interval for r.jk - r.jh: 0.1510 0.3855  
 Null hypothesis rejected (Interval does not include 0)

hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) using average correlation  
 z = 4.4678, p-value = 0.0000  
 Null hypothesis rejected

zou2007: Zou's (2007) confidence interval  
 95% confidence interval for r.jk - r.jh: 0.0676 0.1802  
 Null hypothesis rejected (Interval does not include 0)

Fisher's z-test ( $z = 4.4678$ ,  $p < 0.001$ ) indicates that the correlation of overall instructor with the traditional pedagogy subscale ( $r = 0.80$ ) is higher than its correlation with the socially responsive pedagogy subscale ( $r = 0.67$ ).

```
cocor::cocor(formula = ~Valued + OvInstructor | SRPed + OvInstructor, data = big)
```

Results of a comparison of two overlapping correlations based on dependent groups

Comparison between r.jk (OvInstructor, Valued) = 0.6338 and r.jh (OvInstructor, SRPed) = 0.6717  
 Difference: r.jk - r.jh = -0.0379  
 Related correlation: r.kh = 0.5624  
 Data: big: j = OvInstructor, k = Valued, h = SRPed

Group size: n = 299  
Null hypothesis: r.jk is equal to r.jh  
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)  
Alpha: 0.05

pearson1898: Pearson and Filon's z (1898)  
z = -1.0091, p-value = 0.3129  
Null hypothesis retained

hotelling1940: Hotelling's t (1940)  
t = -1.0355, df = 296, p-value = 0.3013  
Null hypothesis retained

williams1959: Williams' t (1959)  
t = -1.0071, df = 296, p-value = 0.3147  
Null hypothesis retained

olkin1967: Olkin's z (1967)  
z = -1.0091, p-value = 0.3129  
Null hypothesis retained

dunn1969: Dunn and Clark's z (1969)  
z = -1.0062, p-value = 0.3143  
Null hypothesis retained

hendrickson1970: Hendrickson, Stanley, and Hills' (1970) modification of Williams' t (1959)  
t = -1.0355, df = 296, p-value = 0.3013  
Null hypothesis retained

steiger1980: Steiger's (1980) modification of Dunn and Clark's z (1969) using average correlation  
z = -1.0060, p-value = 0.3144  
Null hypothesis retained

meng1992: Meng, Rosenthal, and Rubin's z (1992)  
z = -1.0058, p-value = 0.3145  
Null hypothesis retained  
95% confidence interval for r.jk - r.jh: -0.1948 0.0627  
Null hypothesis retained (Interval includes 0)

hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) using average correlation  
z = -1.0058, p-value = 0.3145  
Null hypothesis retained

zou2007: Zou's (2007) confidence interval  
95% confidence interval for r.jk - r.jh: -0.1129 0.0360  
Null hypothesis retained (Interval includes 0)

Fisher's z-test ( $z = -1.006$ ,  $p = 0.315$ ) indicates that the correlation of overall instructor with the

valued subscale ( $r = 0.4$ ) is is not statistically significantly different than its correlation with the socially responsive pedagogy subscale ( $r = 0.67$ ).

#### 4.7.4 With at least three variables, evaluate the degree to which the instrument demonstrates incremental validity (this should involve two regression equations and their statistical comparison)

Playing around with these variables, let's presume our outcome of interested is the student's *valuation of the class* (i.e., the "Valued by the Student variable") and we usually predict it through traditional pedagogy. What does SRPed contribute over-and-above?

*Please understand, that we would normally have a more robust dataset with other indicators – maybe predicting students grades?*

*Also, we are completely ignoring the multi-level nature of this data. The published manuscript takes a multi-level approach to analyzing the data and my lessons on multi-level modeling address this as well.*

```
big <- na.omit(big) #included b/c there was uneven missingness and the subsequent comparison
Step1 <- lm(Valued ~ TradPed, data = big)
Step2 <- lm(Valued ~ TradPed + SRPed, data = big)
summary(Step1)
```

Call:

```
lm(formula = Valued ~ TradPed, data = big)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.43330	-0.25471	0.04673	0.25388	1.79522

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )							
(Intercept)	1.67482	0.18581	9.014	<0.0000000000000002 ***							
TradPed	0.61426	0.04191	14.656	<0.0000000000000002 ***							
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

Residual standard error: 0.4274 on 213 degrees of freedom

Multiple R-squared: 0.5021, Adjusted R-squared: 0.4998

F-statistic: 214.8 on 1 and 213 DF, p-value: < 0.0000000000000022

```
summary(Step2)
```

Call:

```
lm(formula = Valued ~ TradPed + SRPed, data = big)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.39671	-0.22675	0.03228	0.24841	1.71917

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.44912	0.26349	5.500	0.000000109 ***
TradPed	0.56933	0.05602	10.162 < 0.0000000000000002 ***	
SRPed	0.09116	0.07554	1.207	0.229

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.427 on 212 degrees of freedom

Multiple R-squared: 0.5055, Adjusted R-squared: 0.5009

F-statistic: 108.4 on 2 and 212 DF, p-value: < 0.0000000000000022

In the first step we see that traditional pedagogy had a statistically significant effect on the valued dimension  $B = 0.614, p < 0.001$ . This model accounted for 50% of variance.

In the second step, socially responsive pedagogy was not a statistically significant predictor, over and above traditional pedagogy  $B = 0.091, p = 0.228$ . This model accounted for 51% of variance.

We can formally compare these two models with the *anova()* function in base R.

```
anova(Step1, Step2)
```

Analysis of Variance Table

Model 1: Valued ~ TradPed					
Model 2: Valued ~ TradPed + SRPed					
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	213	38.918			
2	212	38.652	1	0.26554	1.4564 0.2288

We see that socially responsive pedagogy adds only a non-significant proportion of variance over traditional pedagogy ( $F[1, 212] = 38.652, p = 0.229$ ).



# Chapter 5

# Reliability

## Screencasted Lecture Link

The focus of this lecture is the assessment of reliability. We start by defining *classical test theory* and examining several forms of reliability. While the majority of our time is spent considering estimates of internal consistency, we also review retest reliability and inter-rater reliability.

### 5.1 Navigating this Lesson

There is one hour and twenty minutes of lecture. If you work through the materials with me, plan for an additional hour.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's introduction

#### 5.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Define “reliability.”
- Identify broad classes of reliability.
- Interpret reliability coefficients.
- Describe the strengths and limitations of the alpha coefficient.

#### 5.1.2 Planning for Practice

In each of these lessons, I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The practice problems are the start of a larger project that spans multiple lessons. Therefore, if possible, please use a dataset that has item-level data for which there is a theorized total scale score as well as two or more subscales (three subscales is ideal). With each of these options I encourage you to:

- Format (i.e., rescore, if necessary) a dataset so that it is possible to calculate estimates of internal consistency
- Calculate and report the alpha coefficient for a total scale scores and subscales (if the scale has them)
- Calculate and report  $\omega_t$  and  $\omega_h$ . With these two determine what proportion of the variance is due to all the factors, error, and  $g$ .
- Calculate total and subscale scores.
- Describe other reliability estimates that would be appropriate for the measure you are evaluating.

### 5.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (and linked, when possible) in the text with complete citations in the reference list.

- Jhangiani, R. S., Chiang, I.-C. A., Cuttler, C., & Leighton, D. C. (2019). Reliability and Validity. In *Research Methods in Psychology*. <https://doi.org/10.17605/OSF.IO/HF7DQ>
- Revelle, W., & Condon, D. M. (2019a). Reliability from to : A tutorial. Psychological Assessment. <https://doi.org/10.1037/pas0000754>
  - A full-text preprint is available [here](#).
- Revelle, W., & Condon, D. M. (2019b). Reliability from to : A tutorial. Online supplement. Psychological Assessment. <https://doi.org/10.1037/pas0000754>
- Revelle, William. (n.d.). Reliability. In An introduction to psychometric theory with applications in R. Retrieved from <http://www.personality-project.org/dev/r/book/#chapter7>
  - All three documents provide a practical integration of conceptual and mechanical.
- Szymanski, D. M., & Bissonette, D. (2020). Perceptions of the LGBTQ College Campus Climate Scale: Development and psychometric evaluation. *Journal of Homosexuality*, 67(10), 1412–1428. <https://doi.org/10.1080/00918369.2019.1591788>
  - The research vignette for this lesson.

### 5.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(psych)){install.packages('psych')}
# if(!require(tidyverse)){install.packages('tidyverse')}
# if(!require(MASS)){install.packages('MASS')}
# if(!require(sjstats)){install.packages('sjstats')}
# if(!require(apaTables)){install.packages('apaTables')}
# if(!require(qualtrics)){install.packages('qualtrics')}
```

## 5.2 Defining Reliability

### 5.2.1 Begins with Classical Test Theory (CTT)

CTT is based on Spearman's (1904) *true-score model* where:

- an observed score consists of two components – a true component and an error component
- $X = T + E$ 
  - $X$  = the fallible, observed/manifest score, obtained under ideal or perfect conditions of measurement (these conditions never exist);
  - $T$  = the true/latent score (that will likely remain unknown); and
  - $E$  = random error
- In CTT, we assume that the traits measured are constant and the errors random.
  - Therefore, the mean of measurement errors for any individual (upon numerous repeated testings) would be zero.
- That said, in CTT, the true score would be equal to the mean of the observed scores over an indefinite number of repeated measures.
  - Caveat: this is based on the assumption that when individuals are repeatedly measured, their true scores remain unchanged.
- In classic test theory, true score can be estimated over multiple trials. However, if errors are systematically biased, the true score will remain unknown.

### 5.2.2 Why are we concerned with reliability? Error!

Measurements are imperfect and every observation has some unknown amount of error associated with it. There are two components in error:

- **random/unsystematic:** varies in unpredictable and inconsistent ways upon repeated measurements; sources are unknown
- **systematic:** recurs upon repeated measurements reflecting situational or individual effects that, theoretically, could be specified.

Correlations are attenuated (i.e., smaller than) from the true correlation if the observations contain error. Knowing the reliability of an instruments allows us to:

- estimate the degree to which measured at one time and place with one instrument predict scores at another time and/or place and perhaps measured with a different instrument
- estimate the consistency of scores
- estimate "...the degree to which test scores are free from errors of measurement" (APA, 1985, p. 19)

Figure 7.1a in [Revelle's chapter](#) illustrates the *attenuation* of the correlation between the variables  $p$  and  $q$  as a function of reliability.

- circles (latent variables [LV]) represent the *true score*
- observed/measured/manifest variables are represented by squares and each has an associated error; not illustrated are the *random* and *systematic* components of error
- a true score is composed of a measured variable and its error
- the relationship between the true scores would be stronger than the one between the measured variables
- moving to 7.1b in Revelle's chapter, the correlation between LV  $p$  and the observed '' can be estimated from the correlation of  $p'$  with a parallel test (this is the reliability piece)

Figure 7.2 in Revelle's Chapter 7 [] illustrates the conceptual effect of reliability on the estimation of a true score. The figure is meant to demonstrate that when error variances are small and reliability is greater, the variance of the true scores more closely approximates that of observed scores.

### 5.2.3 The Reliability Coefficient

The symbol for reliability,  $r_{xx}$ , sums up the big-picture definition that reliability is the correlation of a measure with itself. There are a number of ways to think about it:

- a "theoretical validity" of a measure because it refers to a relationship between observed scores and scores on a latent variable or construct,
- represents the fraction of an observed score variance that is not error,
- ranges from 0-1
  - 1, when all observed variance is due to true-score variance; there are no random errors,
  - 0, when all observed variance is due to random errors of measurement,
- represents the squared correlation between observed scores and true scores,
- the ratio between true-score variance and observed-score variance (for a formulaic rendition see [[Pedhazur and Schmelkin, 1991](#)]),

$$r_{xt}^2 = r_{xx} = \frac{\sigma_t^2}{\sigma_x^2}$$

where  $r_{xt}^2$  is the proportion of variance between observed scores ( $t + e$ ) and true scores ( $t$ ); its square root is the correlation

$r_{xx}$  is the reliability of a measure

$\sigma_t^2$  is the variance of true scores

$\sigma_x^2$  is the variance of observed scores

- The reliability coefficient is interpreted as the proportion of systematic variance in the observed score.
  - .8 means that 80% of the variance of the observed scores is systematic;
  - .2 (e.g., 1.00 - .8) is the proportion of variance due to random errors;
  - the reliability coefficient is population specific.

To restate the first portion of the formula: although reliability is expressed as a correlation between observed scores, it is also the ratio of reliable variance to total variance.

## 5.3 Research Vignette

The research vignette for this lesson is the development and psychometric evaluation of the Perceptions of the LGBTQ College Campus Climate Scale [Szymanski and Bissonette, 2020]. The scale is six items with responses rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores indicate more negative perceptions of the LGBTQ campus climate. Szymanski and Bissonette [2020] have suggested that the psychometric evaluation supports using the scale in its entirety or as subscales. Each item is listed below with its variable name in parentheses:

- College response to LGBTQ students:
  - My university/college is cold and uncaring toward LGBTQ students. (cold)
  - My university/college is unresponsive to the needs of LGBTQ students. (unresponsive)
  - My university/college provides a supportive environment for LGBTQ students. (unsupportive)
- \* this item must be reverse-scored
- LGBTQ Stigma:
  - Negative attitudes toward LGBTQ persons are openly expressed on my university/college campus. (negative)
  - Heterosexism, homophobia, biphobia, transphobia, and cissexism are visible on my university/college campus. (heterosexism)
  - LGBTQ students are harassed on my university/college campus. (harassed)

A preprint of the article is available at ResearchGate. Below is the script for simulating item-level data from the factor loadings, means, and sample size presented in the published article.

Because data is collected at the item level (and I want this resource to be as practical as possible, I have simulated the data for each of the scales at the item level).

Simulating the data involved using factor loadings, means, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

Five additional scales were reported in the Szymanski and Bissonette article [2020]. Unfortunately, I could not locate factor loadings for all of them; and in two cases, I used estimates from a more recent psychometric analysis. When the individual item and their factor loadings are known, I assigned names based on item content (e.g., “lo\_energy”) rather than using item numbers (e.g., “PHQ4”). When I am doing psychometric analyses, I prefer item-level names so that I can quickly see (without having to look up the item names) how the items are behaving. While the focus of this series of chapters is on the LGBTQ Campus Climate scale, this simulated data might be useful to you in one or more of the suggestions for practice (e.g., examining the psychometric characteristics of one or the other scales). The scales, their original citation, and information about how I simulated data for each are listed below.

- **Sexual Orientation-Based Campus Victimization Scale** [Herek, 1993] is a 9-item item scale with Likert scaling ranging from 0 (*never*) to 3 (*two or more times*). Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.

- **College Satisfaction Scale** [Helm et al., 1998] is a 5-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores represent greater college satisfaction. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.
- **Institutional and Goals Commitment** [Pascarella and Terenzini, 1980] is a 6-item subscale from a 35-item measure assessing academic/social integration and institutional/goal commitment (5 subscales total). The measure had with Likert scaling ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Higher scores on the institutional and goals commitment subscale indicate greater intentions to persist in college. Data were simulated using factor loadings in the source article.
- **GAD-7** [Spitzer et al., 2006] is a 7-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate more anxiety. I simulated data by estimating factor loadings from Brattmyr et al. [2022].
- **PHQ-9** [Kroenke et al., 2001] is a 9-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate higher levels of depression. I simulated data by estimating factor loadings from Brattmyr et al. [2022].

#Entering the intercorrelations, means, and standard deviations from the journal article

```
Szymanski_generating_model <- '
  #measurement model
  CollegeResponse  =~ .88*cold + .73*unresponsive + .73*supportive
  Stigma =~ .86*negative + .76*heterosexism + .71*harassed
  Victimization =~ .8*Vic1 + .8*Vic2 + .8*Vic3 + .8*Vic4 + .8*Vic5 + .8*Vic6 + .8*Vic7 + .8*Vic8 + .8*Vic9
  CollSat =~ .8*Sat1 + .8*Sat2 + .8*Sat3 + .8*Sat4 + .8*Sat5
  Persistence =~ .69*graduation_importance + .63*right_decision + .62*will_register + .59*will_change
  Anxiety =~ .851*nervous + .887*worry_control + .894*much_worry + 674*cant_relax + .484*depression
  Depression =~ .798*anhedonia + .425*down + .591*sleep + .913*lo_energy + .441*appetite

  #Means
  CollegeResponse ~ 2.71*1
  Stigma ~3.61*1
  Victimization ~ 0.11*1
  CollSat ~ 5.61*1
  Persistence ~ 4.41*1
  Anxiety ~ 1.45*1
  Depression ~1.29*1

  #Correlations
  CollegeResponse ~~ .58*Stigma
  CollegeResponse ~~ -.25*Victimization
  CollegeResponse ~~ -.59*CollSat
  CollegeResponse ~~ -.29*Persistence
  CollegeResponse ~~ .17*Anxiety
  CollegeResponse ~~ .18*Depression
```

```

Stigma ~~ .37*Victimization
Stigma ~~ -.41*CollSat
Stigma ~~ -.19*Persistence
Stigma ~~ .27*Anxiety
Stigma ~~ .24*Depression

Victimization ~~ -.22*CollSat
Victimization ~~ -.04*Persistence
Victimization ~~ .23*Anxiety
Victimization ~~ .21*Depression

CollSat ~~ .53*Persistence
CollSat ~~ -.29*Anxiety
CollSat ~~ -.32*Depression

Persistence ~~ -.22*Anxiety
Persistence ~~ -.26*Depression

Anxiety ~~ .76*Depression
'

set.seed(240218)
dfSzy <- lavaan::simulateData(model = Szymanski_generating_model,
                               model.type = "sem",
                               meanstructure = T,
                               sample.nobs=646,
                               standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(dfSzy))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 6 are the Perceptions of LGBTQ Campus Climate Scale
#Rows 7 thru 15 are the Sexual Orientation-Based Campus Victimization Scale
#Rows 16 thru 20 are the College Satisfaction Scale
#Rows 21 thru 26 are the Institutional and Goals Commitment Scale
#Rows 27 thru 33 are the GAD7
#Rows 34 thru 42 are the PHQ9

for(i in 1:ncol(dfSzy)){
  if(i >= 1 & i <= 6){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
  }
  if(i >= 7 & i <= 15){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
  }
  if(i >= 16 & i <= 20){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 10))
  }
}

```

```

dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
}
if(i >= 21 & i <= 26){
dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 5))
}
if(i >= 27 & i <= 33){
dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
}
if(i >= 34 & i <= 42){
dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
}
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
dfSzy <- dfSzy %>% round(0)

#quick check of my work
#psych::describe(dfSzy)

#Reversing the supportive item on the Perceptions of LGBTQ Campus Climate Scale so that the exercises
dfSzy <- dfSzy %>%
  dplyr::mutate(supportiveNR = 8 - supportive)

#Reversing three items on the Institutional and Goals Commitments scale so that the exercises
dfSzy <- dfSzy %>%
  dplyr::mutate(not_graduateNR = 8 - not_graduate)%>%
  dplyr::mutate(undecidedNR = 8 - undecided)%>%
  dplyr::mutate(grades_unimportantNR = 8 - grades_unimportant)

dfSzy <- dplyr::select(dfSzy, -c(supportive, not_graduate, undecided, grades_unimportant))

```

The optional script below will let you save the simulated data to your computing environment as either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(dfSzy, 'SzyDF.rds') bring back the simulated dat from
# an .rds file dfSzy <- readRDS('SzyDF.rds')

```

```

# write the simulated data as a .csv write.table(dfSzy,
# file='SzyDF.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file dfSzy <-
# read.csv('SzyDF.csv', header = TRUE)

```

If we look at the information about this particular scale, we recognize that the *supportive* item is scaled in the opposite direction of the rest of the items. That is, a higher score on *supportive* would indicate a positive perception of the campus climate for LGBTQ individuals, whereas higher scores on the remaining items indicate a more negative perception. Before moving forward, we must reverse score this item.

In this psychometrics example I have given my variables one-word names that represent each item. Many researchers (including myself) will often give variable names that are alpha numerical: LGBTQ1, LGBTQ2, LGBTQn. In the psychometric evaluations, the one-word names may be useful shortcuts as one begins to understand the inter-item relations.

In reverse-scoring the *supportive* item, I will rename it “*unsupportive*” as an indication of its reversed direction.

```
library(tidyverse)

dfSzy <- dfSzy %>%
  dplyr::mutate(unsupportive = 8 - supportiveNR) #when reverse-coding, subtract the variable

# When unhashtagged, this code provides item-level descriptive
# statistics psych::describe(dfSzy)
```

Next, I will create dfs that each contain the items of the total and subscales. These will be useful in the reliability estimates that follow.

Note that I am adding “T1” (time 1) to the end of the variable names. Later in the lesson when we evaluate test-retest reliability, we will simulate and add a “T2.”

```
LGBTQT1 <- dplyr::select(dfSzy, cold, unresponsive, unsupportive, negative,
  heterosexism, harassed)
ResponseT1 <- dplyr::select(dfSzy, cold, unresponsive, unsupportive)
StigmaT1 <- dplyr::select(dfSzy, negative, heterosexism, harassed)
```

As we move into the lecture, allow me to provide a content advisory. Individuals who hold LGBTQIA+ identities are frequently the recipients of discrimination and harassment. If you are curious about why these items are considered to be stigmatizing or non-responsive, please do not ask a member of the LGBTQIA+ community to explain it to you; it is not their job to educate others on discrimination, harassment, and microaggressions. Rather, please read the article in its entirety. Additionally, resources such as [The Trevor Project](#), [GLSEN](#), and [Campus Pride](#) are credible sources of information for learning more.

## 5.4 A Parade of Reliability Coefficients

While I cluster the reliability coefficients into large groups, please understand that these are somewhat overlapping.

Table 1 in Revelle and Condon’s [2019b] article provides a summary of the type of reliability tested, the findings, and the function used in the *psych* package.

### 5.4.1 Reliability Options for a Single Administration

If reliability is defined as the correlation between a test and a test just like it, how do we estimate the reliability of a single test, given only one time [Revelle and Condon, 2019b]? It may help to keep in mind that reliability is the ratio of true score variance to test score variance (or  $1 - \text{the ratio of error variance}$ ). Thus, the goal is to estimate the amount of error variance in the test. In this case we can investigate:

- a correlation between two random parts of the test
- internal consistency
- the internal structure of the test

#### 5.4.1.1 Split-half reliability

*Split-half reliability* is splitting a test into two random halves, correlating the two halves, and adjusting the correlation with the *Spearman-Brown* prophecy formula. Abundant formulaic detail in Revelle's Chapter 7/Reliability [].

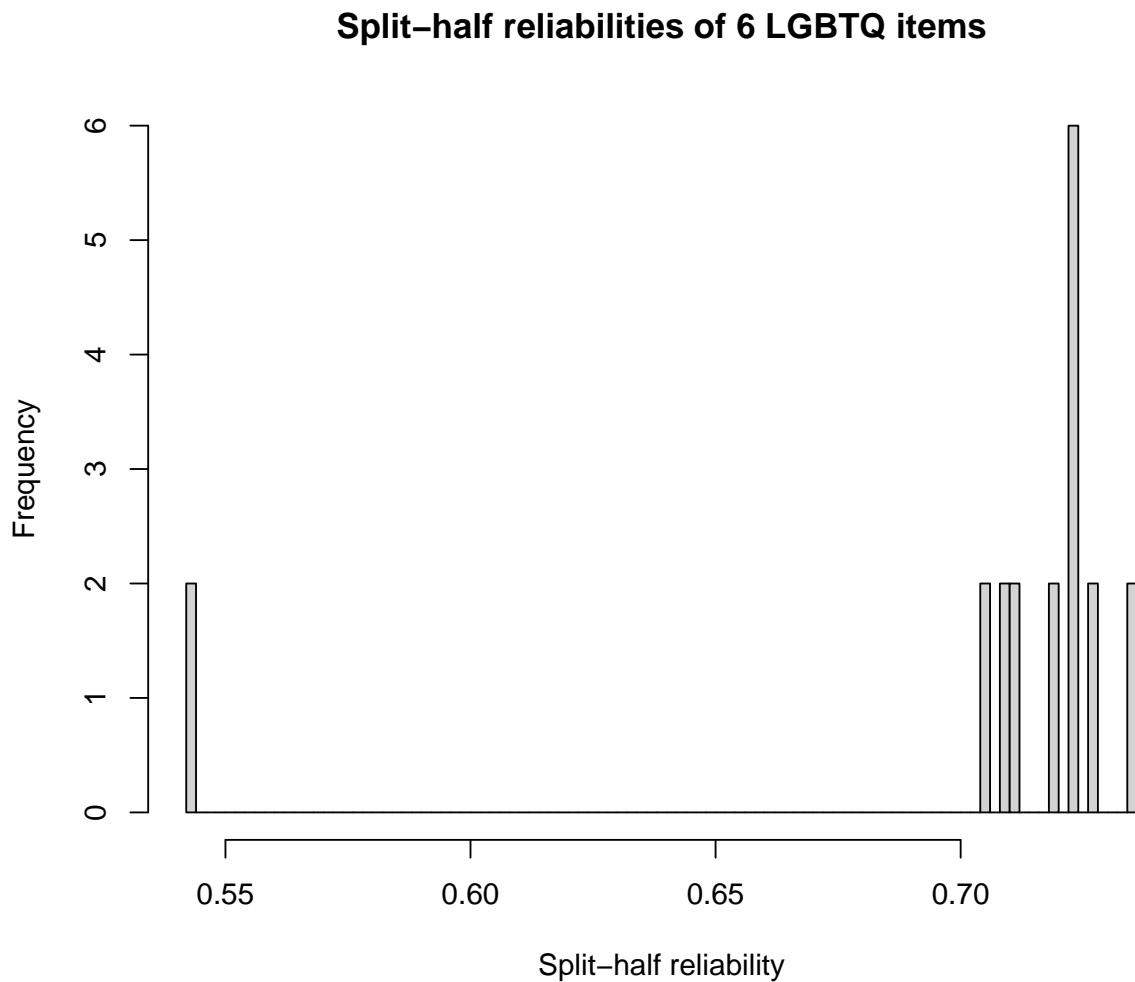
An important question to split-half is “How to split?” Revelle terms it a “combinatorially difficult problem.” There are 126 possible splits for a 10 item scale; 6,345 possible splits for a 16 item scale; and over 4.5 billion for a 36 item scale! The *psych* package’s *splitHalf()* function will try all possible splits for scales of up to 16 items, then sample 10,000 splits for scales longer than that.

```
split <- psych::splitHalf(LGBTQT1, raw = TRUE, brute = TRUE)
split #show the results of the analysis
```

```
Split half reliabilities
Call: psych::splitHalf(r = LGBTQT1, raw = TRUE, brute = TRUE)

Maximum split half reliability (lambda 4) = 0.73
Guttman lambda 6 = 0.68
Average split half reliability = 0.7
Guttman lambda 3 (alpha) = 0.7
Guttman lambda 2 = 0.71
Minimum split half reliability (beta) = 0.54
Average interitem r = 0.28 with median = 0.25
                                         2.5% 50% 97.5%
Quantiles of split half reliability = 0.54 0.72 0.73
```

```
hist(split$raw, breaks = 101, xlab = "Split-half reliability", main = "Split-half reliabilities")
```



Results of the split-half can provide some indication of whether not the scale is unidimensional.

In this case, the maximum reliability coefficient is 0.73, the average 0.70, and the lowest is 0.54. Similarly, we can examine the quantiles: 0.54, 0.72, 0.73.

The split-half output also includes the classic Cronbach's (1951) alpha coefficient (0.70; aka Guttman lambda 3) and average interitem correlations (0.25). The figure plots the frequencies of the reliability coefficient values.

While I did not find guidelines on what constitutes a “high enough lower bound” to establish homogeneity, Revelle suggested that a scale with 0.85, 0.80, and 0.65 had “strong evidence for a relatively homogeneous scale.” When the values were 0.81, 0.73, 0.42, Revelle indicated that there was “strong evidence for non-homogeneity” [Revelle and Condon, 2019a, p. 11]. In making this declaration, Revelle was also looking at the strength of the inter-item correlation and for a rather tight, bell-shaped, distribution at the higher ( $> 0.73$ ) end of the figure.

What happens when we examine the split-half estimates of the subscales? With only three items, there’s not much of a split and so the associated histogram will not be helpful.

```
splitRx <- psych::splitHalf(ResponseT1, raw = TRUE, brute = TRUE)
splitRx #show the results of the analysis
```

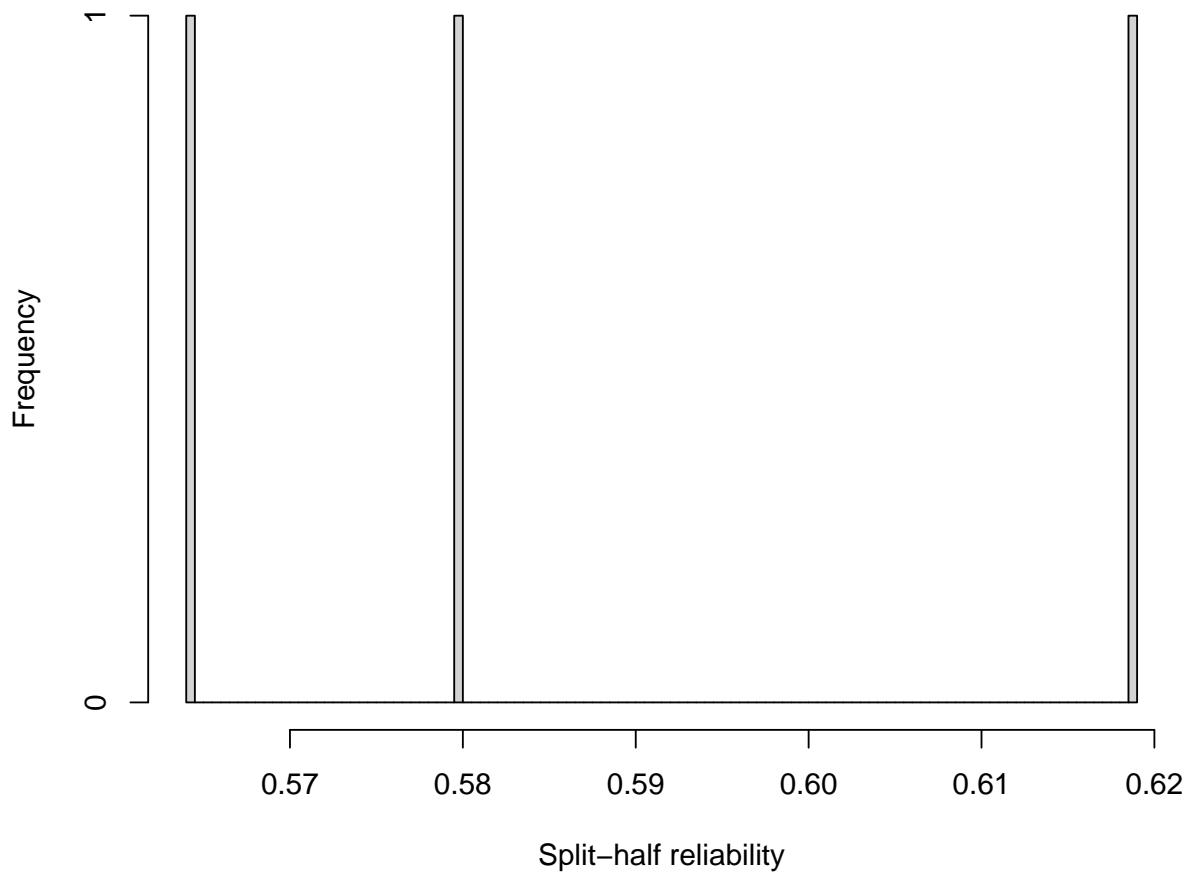
#### Split half reliabilities

```
Call: psych::splitHalf(r = ResponseT1, raw = TRUE, brute = TRUE)
```

```
Maximum split half reliability (lambda 4) = 0.62
Guttman lambda 6                      = 0.57
Average split half reliability          = 0.59
Guttman lambda 3 (alpha)                = 0.66
Guttman lambda 2                      = 0.66
Minimum split half reliability (beta)   = 0.56
Average interitem r = 0.39 with median = 0.4
                                         2.5% 50% 97.5%
Quantiles of split half reliability    = 0.57 0.58 0.62
```

```
hist(splitRx$raw, breaks = 101, xlab = "Split-half reliability", main = "Split-half reliabilit
```

### Split-half reliabilities of 3 items of the College Response subscale



The alpha is 0.66. The range of splits for max, ave, and low are 0.62, 0.59, and 0.55 and the quantiles are 0.57, 0.58, 0.62. The inter-item correlations have an average of 0.40.

Let's look at the split-half reliability coefficients for the Stigma subscale.

```
splitSt <- psych::splitHalf(StigmaT1, raw = TRUE, brute = TRUE)
splitSt #show the results of the analysis
```

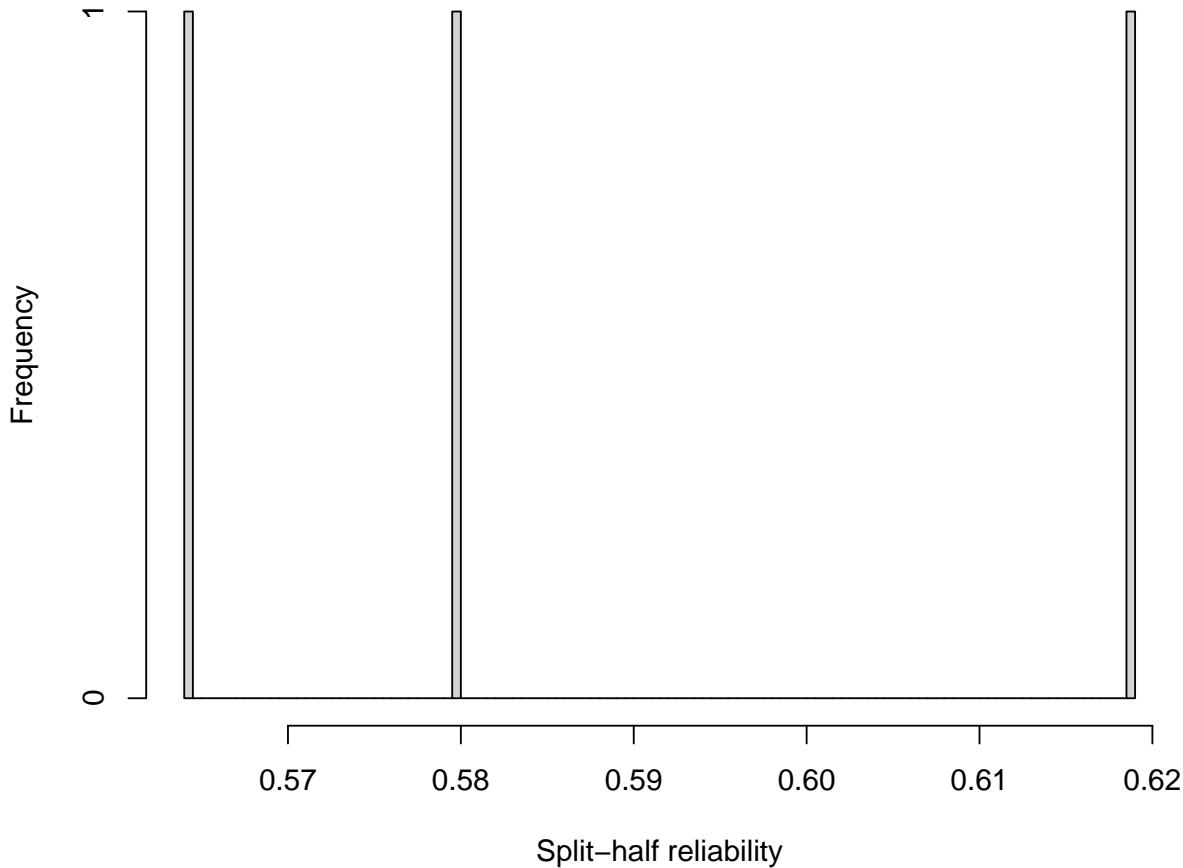
```
Split half reliabilities
Call: psych::splitHalf(r = StigmaT1, raw = TRUE, brute = TRUE)

Maximum split half reliability (lambda 4) =  0.56
Guttman lambda 6                      =  0.53
Average split half reliability        =  0.56
Guttman lambda 3 (alpha)              =  0.63
Guttman lambda 2                      =  0.63
Minimum split half reliability (beta) =  0.55
Average interitem r =  0.36  with median =  0.36
```

Quantiles of split half reliability      =    2.5% 50% 97.5%  
    =    0.55 0.56 0.56

```
hist(splitRx$raw, breaks = 101, xlab = "Split-half reliability", main = "Split-half reliabiliti
```

### Split-half reliabilities of 3 items of the Stigma subscale



The alpha of this subscale is lower than the total scale score ( $\alpha = 0.60$ ). The maximum, average, and minimum split-half reliabilities were 0.56, 0.56, and 0.55; quantiles were at 0.55, 0.56, and 0.56. The average interitem correlation was 0.36.

Because the alpha coefficient can be defined as the “average of all possible split-half coefficients” for the groups tested, it is common for researchers to not provide split-half results in their papers – this is true for our research vignette. I continue to teach the split-half because it can be a stepping stone in the conceptualization of internal consistency as an estimate of reliability.

#### 5.4.1.2 Alpha coefficients

The most common methods to assess internal consistency are the *KR20* (for dichotomous items) and  $\alpha$  (for Likert scaling); alpha has an alias,  $\lambda_3$  (i.e., the Guttman lambda 3).

Alpha and the Guttman 3 (used for scales with Likert-type scaling) may be thought of as:

- a function of the number of items and the average correlation between the items
- the correlation of a test with a non-existent test just like it
- average of all possible split-half coefficients for the groups tested

Although the *psych* package has an incredible and thorough *alpha()* function, Revelle is not a fan of alpha. In fact, his alpha function reports a 95% CI around alpha as well as bootstrapped alpha results.

Let's grab alpha coefficients for our total and subscales.

```
psych::alpha(LGBTQT1)
```

Reliability analysis

Call: psych::alpha(x = LGBTQT1)

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r	
	0.7	0.7	0.68		0.28	2.4	0.018	4	0.63	0.25

95% confidence boundaries

	lower	alpha	upper
--	-------	-------	-------

Feldt	0.66	0.7	0.74
-------	------	-----	------

Duhachek	0.66	0.7	0.74
----------	------	-----	------

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
cold	0.64	0.64	0.61		0.27	1.8	0.022	0.0066	0.22
unresponsive	0.66	0.66	0.63		0.28	2.0	0.021	0.0073	0.25
unsupportedive	0.67	0.67	0.63		0.29	2.0	0.021	0.0058	0.25
negative	0.66	0.66	0.63		0.28	2.0	0.021	0.0084	0.25
heterosexism	0.66	0.66	0.63		0.28	2.0	0.021	0.0087	0.25
harassed	0.67	0.67	0.64		0.29	2.0	0.021	0.0078	0.25

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
cold	646	0.68	0.68	0.59	0.49	4.1	1.03
unresponsive	646	0.63	0.63	0.51	0.43	4.3	0.99
unsupportedive	646	0.62	0.62	0.51	0.42	3.7	0.98
negative	646	0.64	0.63	0.51	0.42	4.0	1.04
heterosexism	646	0.61	0.63	0.51	0.43	4.0	0.90
harassed	646	0.63	0.61	0.49	0.41	3.9	1.07

Non missing response frequency for each item

	1	2	3	4	5	6	7	miss
cold	0.00	0.04	0.22	0.40	0.23	0.09	0.00	0

unresponsive	0.00	0.03	0.17	0.37	0.33	0.09	0.01	0
unsupportive	0.01	0.07	0.35	0.37	0.17	0.02	0.01	0
negative	0.01	0.07	0.23	0.39	0.24	0.05	0.00	0
heterosexism	0.00	0.03	0.24	0.43	0.26	0.03	0.00	0
harassed	0.01	0.07	0.27	0.37	0.22	0.05	0.01	0

The second screen of output shows the information we are interested in:

- **raw\_alpha**, .70 is based on the covariances
- **std.alpha**, .70 is based on correlations
- **average\_r**, .28 is the average inter-item correlation (i.e., all possible pairwise combinations of items)

```
psych::alpha(ResponseT1)
```

Reliability analysis

Call: psych::alpha(x = ResponseT1)

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.66	0.66	0.57	0.39	1.9	0.023	4	0.77	0.4

95% confidence boundaries

lower alpha upper

Feldt	0.61	0.66	0.70
-------	------	------	------

Duhachek	0.62	0.66	0.71
----------	------	------	------

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
cold	0.52	0.52	0.35	0.35	1.1	0.038	NA	0.35	
unresponsive	0.60	0.60	0.42	0.42	1.5	0.032	NA	0.42	
unsupported	0.58	0.58	0.40	0.40	1.4	0.033	NA	0.40	

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
cold	646	0.80	0.79	0.62	0.50	4.1	1.03
unresponsive	646	0.76	0.76	0.55	0.45	4.3	0.99
unsupported	646	0.76	0.77	0.57	0.46	3.7	0.98

Non missing response frequency for each item

	1	2	3	4	5	6	7	miss
cold	0.00	0.04	0.22	0.40	0.23	0.09	0.00	0
unresponsive	0.00	0.03	0.17	0.37	0.33	0.09	0.01	0
unsupported	0.01	0.07	0.35	0.37	0.17	0.02	0.01	0

In the case of the College Response subscale:

- **raw\_alpha**, .66 is based on the covariances
- **std.apha**, .66 is based on correlations
- **average\_r**, .39 is the average interitem correlation

```
psych::alpha(StigmaT1)
```

Reliability analysis

Call: psych::alpha(x = StigmaT1)

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.62	0.63	0.53	0.36	1.7	0.025	4	0.76	0.36

95% confidence boundaries

lower alpha upper

Feldt 0.57 0.62 0.67

Duhachek 0.57 0.62 0.67

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
negative	0.52	0.52	0.35	0.35	1.1	0.038	NA	0.35	
heterosexism	0.53	0.53	0.36	0.36	1.1	0.037	NA	0.36	
harassed	0.53	0.54	0.37	0.37	1.2	0.036	NA	0.37	

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
negative	646	0.77	0.76	0.56	0.44	4.0	1.0
heterosexism	646	0.73	0.76	0.55	0.44	4.0	0.9
harassed	646	0.77	0.75	0.54	0.43	3.9	1.1

Non missing response frequency for each item

	1	2	3	4	5	6	7	miss
negative	0.01	0.07	0.23	0.39	0.24	0.05	0.00	0
heterosexism	0.00	0.03	0.24	0.43	0.26	0.03	0.00	0
harassed	0.01	0.07	0.27	0.37	0.22	0.05	0.01	0

In the case of the Stigma subscale:

- **raw\_alpha**, .62 is based on the covariances
- **std.apha**, .63 is based on correlations
- **average\_r**, .36 is the average interitem correlation

The documentation for this package is incredible. Scrolling down through the description of the *alpha()* function provides a description of these different statistics.

Especially useful are item-level statistics:

- **r.drop** is the corrected item-total correlation (in the next lesson) for this item against the remaining items in the scale
- **mean** and **sd** are the mean and standard deviation of each item across all individuals.

The popularity of alpha emerged when tools available for calculation were less sophisticated; since then we have learned that alpha can be misleading:

- alpha inflates, somewhat artificially, even when inter-item correlations are low.
  - a 14-item scale will have an alpha of at least .70, even if it has two orthogonal (i.e., unrelated) scales [Cortina, 1993],
- alpha assumes a unidimensional factor structure,
- the same alpha can be obtained for dramatically different underlying factor structures (see graphs in Revelle's Chapter 7)

The proper use of alpha requires the following:

- *tau equivalence*, that is, equal covariances with the latent score represented by the test, and
- *unidimensionality*, equal factor loadings on the single factor of the test

When either of these is violated, alpha underestimates reliability and overestimates the fraction of test variance that is associated with the general variance in the test.

Alpha and the split half are *internal consistency* estimates. Moving to *model-based* techniques allows us to take into consideration the factor structure of the scale. In the original article [Szymanski and Bissonette, 2020], results were as follows:

#### 5.4.1.3 Omega

Assessing reliability with *omega* ( $\omega$ ) statistics falls into a larger realm of *composite reliability* where reliability is assessed from a ratio of the variability explained by the items compared with the total variance of the entire scale [McNeish, 2018]. Members of the omega family of reliability estimates come from factor exploratory (i.e., EFA) and confirmatory (i.e., CFA; structural equation modeling [SEM]) factor analytic approaches. This lesson precedes the lessons on CFA and SEM. Therefore, my explanations and demonstrations will be somewhat brief. I intend to revisit omega output in the CFA and SEM lessons and encourage you to review this section now, then return to this section again after learning more about CFA and SEM.

In the context of *psychometrics* it may be useful (albeit an oversimplification) to think of factors as scales/subscales where  $g$  refers to the amount of variance in the *general* factor (or total scale score) and subscales to be items that have something in common that is separate from what is  $g$ .

Model-based estimates examine the correlations or covariances of the items and decompose the test variance into that which is:

- common to all items ( $g$ , a general factor),
- specific to some items ( $f$ , orthogonal group factors), and

- unique to each item (confounding **s** specific, and **e** error variance)

$\omega$  is something of a shapeshifter. In the *psych* package:

- $\omega_t$  represents the total reliability of the test ( $\omega_t$ )
  - In the *psych* package, this is calculated from a bifactor model where there is one general *g* factor (i.e., each item loads on the single general factor), one or more group factors (*f*), and item-specific factors.
- $\omega_h$  extracts a higher-order factor from the correlation matrix of lower level factors, then applies the Schmid and Leiman [1957] transformation to find the general loadings on the original items. Stated another way, it is a measure of the general factor saturation (*g*; the amount of variance attributable to one common factor). The subscript “*h*” acknowledges the hierarchical nature of the approach.
  - the  $\omega_h$  approach is exploratory and defined if there are three or more group factors (with only two group factors, the default is to assume they are equally important, hence the factor loadings of those subscales will be equal)
  - Najera Catalan [Najera Catalan, 2019] suggests that  $\omega_h$  is the best measure of reliability when dealing with multiple dimensions.
- $\omega_g$  is an estimate that uses a bifactor solution via the SEM package *lavaan* and tends to be a larger (because it forces all the cross loadings of lower level factors to be 0)
  - the  $\omega_g$  is confirmatory, requiring the specification of which variables load on each group factor

Two commands in *psych* get us the results:

- *omega()* reports only the EFA solution
- *omegaSem()* reports both EFA and CFA solutions
  - We will use the *omegaSem()* function

Note that in our specification, we indicate there are two factors. We do not tell it what items belong to what factors (think, *subscals*). One test will be to see if the items align with their respective factors.

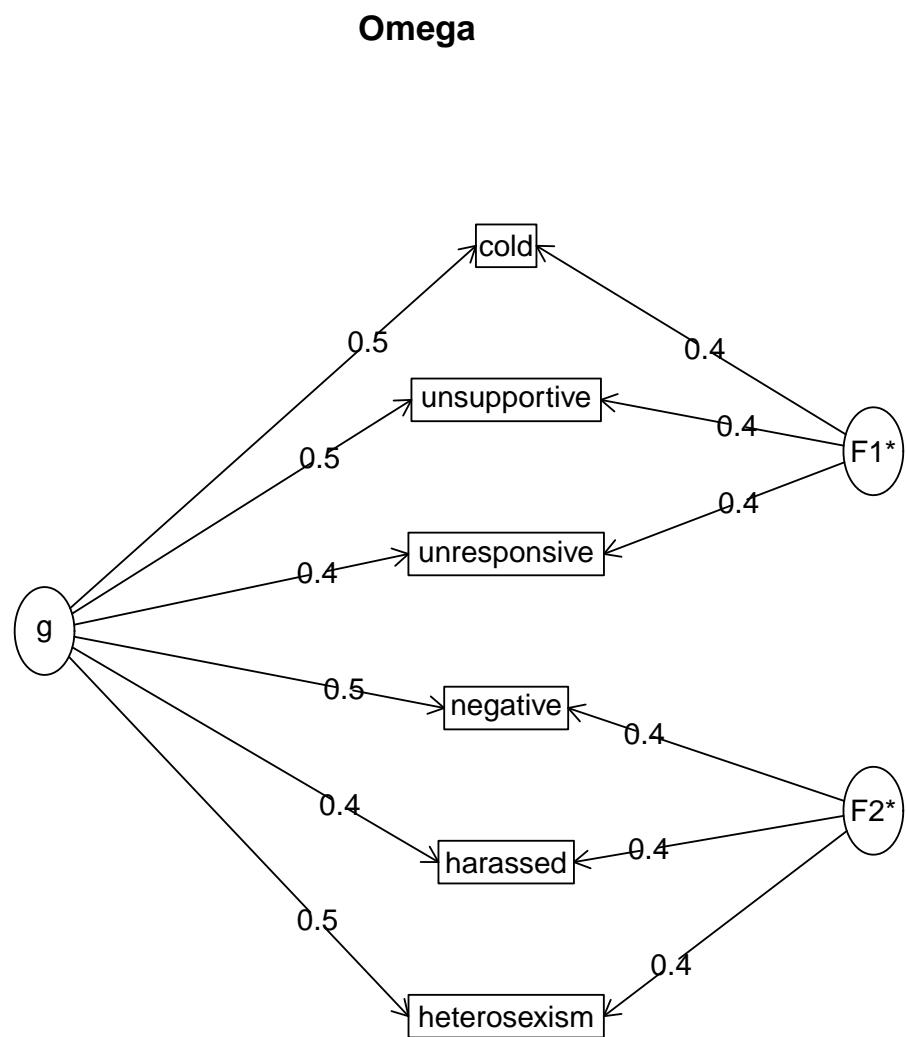
```
psych::omegaSem(LGBTQT1, nfactors = 2)
```

Loading required namespace: GPArotation

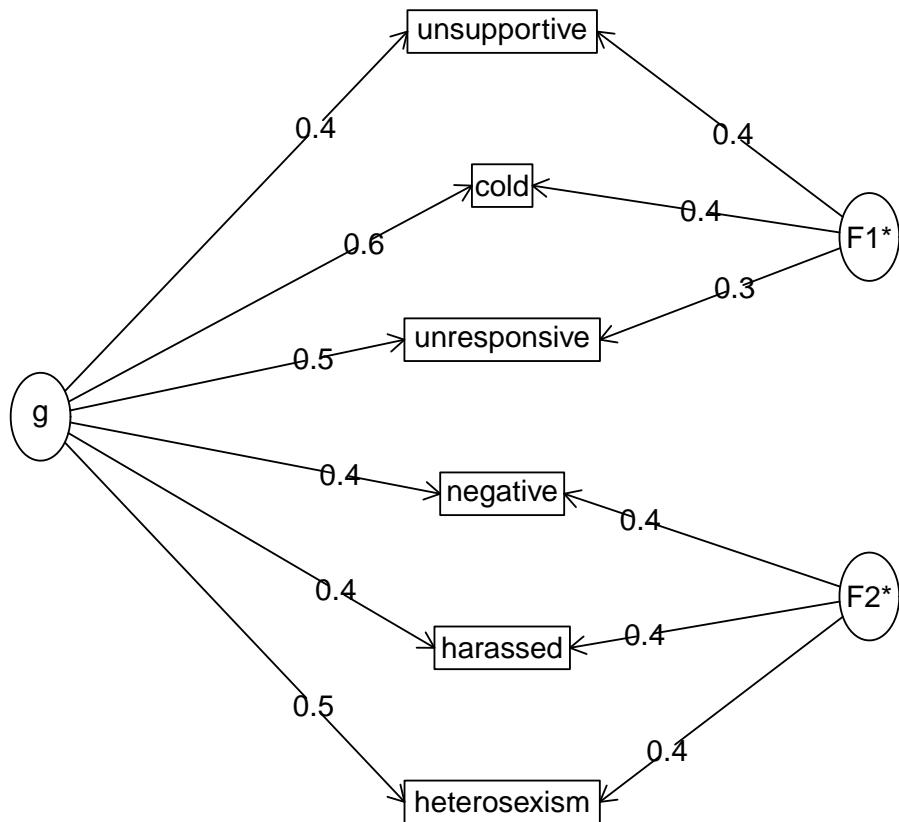
Three factors are required for identification -- general factor loadings set to be equal.  
Proceed with caution.

Think about redoing the analysis with alternative values of the 'option' setting.

Warning in lav\_model\_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WARNING  
Could not compute standard errors! The information matrix could  
not be inverted. This may be a symptom that the model is not  
identified.



### Omega from SEM



```

Call: psych::omegaSem(m = LGBTQT1, nfactors = 2)
Omega
Call: omegah(m = m, nfactors = nfactors, fm = fm, key = key, flip = flip,
  digits = digits, title = title, sl = sl, labels = labels,
  plot = plot, n.obs = n.obs, rotate = rotate, Phi = Phi, option = option,
  covar = covar)

```

Alpha:	0.7
G.6:	0.68
Omega Hierarchical:	0.54
Omega H asymptotic:	0.73
Omega Total	0.74

Schmid Leiman Factor loadings greater than 0.2

	<i>g</i>	<i>F1*</i>	<i>F2*</i>	<i>h2</i>	<i>u2</i>	<i>p2</i>
<i>cold</i>	0.53	0.45		0.49	0.51	0.59
<i>unresponsive</i>	0.45	0.37		0.34	0.66	0.60

unsupportive	0.45	0.41	0.37	0.63	0.55	
negative	0.46		0.40	0.37	0.63	0.58
heterosexism	0.46		0.39	0.36	0.64	0.59
harassed	0.44		0.39	0.35	0.65	0.56

With Sums of squares of:

g	F1*	F2*
1.31	0.51	0.46

general/max 2.59 max/min = 1.1

mean percent general = 0.58 with sd = 0.02 and cv of 0.03

Explained Common Variance of the general factor = 0.58

The degrees of freedom are 4 and the fit is 0

The number of observations was 646 with Chi Square = 2.59 with prob < 0.63

The root mean square of the residuals is 0.01

The df corrected root mean square of the residuals is 0.02

RMSEA index = 0 and the 10 % confidence intervals are 0 0.049

BIC = -23.3

Compare this with the adequacy of just a general factor and no group factors

The degrees of freedom for just the general factor are 9 and the fit is 0.18

The number of observations was 646 with Chi Square = 115.31 with prob < 0.0000000000000000

The root mean square of the residuals is 0.1

The df corrected root mean square of the residuals is 0.13

RMSEA index = 0.135 and the 10 % confidence intervals are 0.114 0.158

BIC = 57.08

Measures of factor score adequacy

	g	F1*	F2*
Correlation of scores with factors	0.74	0.57	0.56
Multiple R square of scores with factors	0.54	0.33	0.31
Minimum correlation of factor score estimates	0.09	-0.34	-0.38

Total, General and Subset omega for each subset

	g	F1*	F2*
Omega total for total scores and subscales	0.74	0.66	0.63
Omega general for total scores and subscales	0.54	0.38	0.36
Omega group for total scores and subscales	0.20	0.28	0.27

The following analyses were done using the lavaan package

Omega Hierarchical from a confirmatory model using sem = 0.54

Omega Total from a confirmatory model using sem = 0.74

With loadings of

	g	F1*	F2*	h2	u2	p2	
cold		0.56	0.42		0.49	0.51	0.64

unresponsive	0.47	0.33	0.34	0.66	0.65
unsupportive	0.44	0.42	0.38	0.62	0.51
negative	0.45		0.42	0.37	0.63
heterosexism	0.46		0.39	0.36	0.64
harassed	0.43		0.40	0.34	0.66
					0.54

With sum of squared loadings of:

g	F1*	F2*
1.32	0.46	0.49

The degrees of freedom of the confirmatory model are 3 and the fit is 2.658827 with p = 0  
 general/max 2.71 max/min = 1.05  
 mean percent general = 0.58 with sd = 0.06 and cv of 0.1  
 Explained Common Variance of the general factor = 0.58

Measures of factor score adequacy

	g	F1*	F2*
Correlation of scores with factors	0.74	0.55	0.58
Multiple R square of scores with factors	0.55	0.31	0.33
Minimum correlation of factor score estimates	0.10	-0.39	-0.34

Total, General and Subset omega for each subset

	g	F1*	F2*
Omega total for total scores and subscales	0.74	0.66	0.63
Omega general for total scores and subscales	0.54	0.41	0.34
Omega group for total scores and subscales	0.20	0.26	0.28

To get the standard sem fit statistics, ask for summary on the fitted object

There is a ton of output! How do we make sense of it?

First, our items aligned perfectly with their respective factors (subscales). That is, it would be problematic if the items switched factors.

Second, we can interpret our results. Like alpha, the omegas range from 0 to 1, where values closer to 1 represent good reliability [Najera Catalan, 2019]. For unidimensional measures,  $\omega_t$  values above 0.80 indicate satisfactory reliability. For multidimensional measures with well-defined dimensions, we strive for  $\omega_h$  values above 0.65 (and  $\omega_t > 0.8$ ). These recommendations are based on a Monte Carlo study that examined a host of reliability indicators and how their values corresponded with accurate predictions of poverty status. With this in mind, let's examine the output related to our simulated research vignette.

Let's start with the output in the lower portion where the values are "from a confirmatory model using sem."

Omega is a reliability estimate for factor analysis that represents the proportion of variance in the LGBTQ scale attributable to common variance rather than error. The omega for the total reliability of the test ( $\omega_t$ ; which included the general factors and the subscale factors) was .74, meaning that 74% of the variance in the total scale is due to the factors and 26% (100% - 74%) is attributable to error.

Omega hierarchical ( $\omega_h$ ) estimates are the proportion of variance in the LGBTQ score attributable to the general factor, which in effect treats the subscales as error.  $\omega_h$  for the the LGBTQ total scale was .54. A quick calculation with  $\omega_h$  (.54) and  $\omega_t$  (.74;  $.54/.74 = .72$ ) lets us know that that 73% of the reliable variance in the LGBTQ total scale is attributable to the general factor.

```
.54/.74
```

```
[1] 0.7297297
```

Amongst the output is the Cronbach's alpha coefficient (.70). Szymanski and Bissonette [2020] did not report omega results; this may be because there were only two subfactors and/or they did not feel like a bifactor analysis would be appropriate. You might notice the lavaan warning indicating that three factors are needed in order to identify the CFA model. There is a longer explanation about factor identification. Stay tuned for CFA models.

#### 5.4.1.4 Some summary statements about reliability from single administrations

- With the exception of the worst split-half reliability and  $\omega_g$  or  $\omega_h$ , all of the reliability estimates are functions of test length and will tend asymptotically towards 1 as the number of items increases
- the omega output provides a great deal more information about reliability than a simple alpha
  - Figure 7.5 in [Revelle's chapter](#) shows four different structural representations of measures that have equal alphas (all .72)
- $\omega_{(h)}$ ,  $\beta$ , and the worst split-half reliability are estimates of the amount of general factor variance in the test scores
- in the case of low general factor saturation, the EFA based  $\omega_{(h)}$  is positively biased, so the CFA-based estimate,  $\omega_{(g)}$ , should be used
- $\omega_{(t)}$  is the model-based estimate of the greatest lower bound of the total reliability of the test; so is the best split-half reliability

Revelle and Condon's [2019a] recommendations to researchers:

- report at least two coefficients (e.g.,  $\omega_{(h)}$  and  $\omega_{(t)}$ ) and discuss why each is appropriate for the inference that is being made,
- report more than "just alpha" unless you can demonstrate that the measure is tau equivalent and unidimensional

#### 5.4.2 Reliability Options for Two or more Administrations

##### 5.4.2.1 Test-retest of total scores

The purpose of test-retest reliability is to understand the stability of the measure over time. With two time points, T1 and T2, the test-retest correlation is an unknown mixture of trait, state, and specific variance, and is a function of the length of time between two measures.

- With two time points we cannot distinguish between trait and state effects, that said
  - we would expect a high degree of stability if the retest is (relatively) immediate
- With three time points we can leverage some SEM tools to distinguish between trait and state components
- A large test-retest correlation over a long period of time indicates temporal stability. Temporal stability is:
  - expected if we are assessing something trait like (e.g., cognitive ability, personality trait)
  - not expected if we are assessing something state like (e.g., emotional state, mood)
  - not expected if there was an intervention (or condition) and the T1 and T2 administrations are part of a pre- and post-test design.

There are some *methodological* concerns about test-retest reliability. For example, owing to memory and learning effects, the average response time to a second administration of identical items takes about 80% the time compared to the first administration.

Szymanski and Bissonette [2020] did not assess retest reliability. We can, though, imagine how this might work. Let's imagine that both waves were taken in the same academic term, approximately two weeks apart.

With both sets of data we need to create scores for the total scale score and the two subscales. We would also need to join the two datasets into a single dataframe.

To demonstrate the retest reliability, I simulated a new dataset with total and subscale scores for our variables for Time 1 and Time 2. This next script is simply that simulation (i.e., you can skip over it). If this were your data, you would have item-level data and need to calculate total and subscale scores (as we did above).

```
SimCor_mu <- c(3.13, 2.68, 3.58, 3.16, 2.66, 2.76)
SimCor_sd <- c(0.82, 1.04, 1.26, 0.83, 1.05, 0.99)
simCor <- matrix(c(1, 0.64, 0.77, 0.44, 0.33, 0.29, 0.64, 1, 0.53, 0.35,
  0.46, 0.34, 0.77, 0.53, 1, 0.27, 0.4, 0.47, 0.44, 0.35, 0.27, 1, 0.63,
  0.62, 0.33, 0.46, 0.4, 0.63, 1, 0.57, 0.29, 0.34, 0.47, 0.62, 0.57,
  1), ncol = 6)
scovMat <- SimCor_sd %*% t(SimCor_sd) * simCor
set.seed(210829)
retest_df <- MASS::mvrnorm(n = 646, mu = SimCor_mu, Sigma = scovMat, empirical = TRUE)
colnames(retest_df) <- c("TotalT1", "ResponseT1", "StigmaT1", "TotalT2",
  "ResponseT2", "StigmaT2")
retest_df <- as.data.frame(retest_df) #converts to a df so we can use in R
library(dplyr)
retest_df <- retest_df %>%
  dplyr::mutate(ID = row_number()) #add ID to each row
retest_df <- retest_df %>%
  dplyr::select(ID, everything()) #moving the ID number to the first column; requires
```

Examining our df, we can see the ID variable and the three sets of scores for each wave of analysis. Now we simply ask for their correlations. There are a number of ways to do this – the *apaTables* package can do the calculations and pop it into a manuscript-ready table.

We won't want the ID variable to be in the table.

```
retest_df2 <- retest_df %>%
  dplyr::select(c(-ID))

apaTables::apa.cor.table(data = retest_df2, landscape = TRUE, table.number = 1,
  filename = "Table_1_Retest.doc")
```

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3	4
1. TotalT1	3.13	0.82				
2. ResponseT1	2.68	1.04	.64** [.59, .68]			
3. StigmaT1	3.58	1.26	.77** [.74, .80]	.53** [.47, .58]		
4. TotalT2	3.16	0.83	.44** [.38, .50]	.35** [.28, .42]	.27** [.20, .34]	
5. ResponseT2	2.66	1.05	.33** [.26, .40]	.46** [.40, .52]	.40** [.33, .46]	.63** [.58, .67]
6. StigmaT2	2.76	0.99	.29** [.22, .36]	.34** [.27, .41]	.47** [.41, .53]	.62** [.57, .67]

.57\*\*  
 [.52, .62]

Note. M and SD are used to represent mean and standard deviation, respectively.  
 Values in square brackets indicate the 95% confidence interval.  
 The confidence interval is a plausible range of population correlations  
 that could have caused the sample correlation (Cumming, 2014).  
 \* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

As expected in this simulation,

- the strongest correlations are within each scale at their respective time, that is:
  - the T1 variables correlate with each other;
  - the T2 variables correlate with each other.
- the next strongest correlations are with the same scale/subscale configuration across time, for example
  - TotalT1 with TotalT2 ( $r = .44, p < 0.01$ )
  - ResponseT1 with ResponseT2 ( $r = .46, p < 0.01$ )
  - StigmaT1 with StigmaT2 ( $r = .47, p < 0.01$ )
- the lowest correlations are different scales at T1 and T2
  - ResponseT1 with StigmaT2 ( $r = .29$ )

The range of retest correlations (e.g., .44 to .47 with  $p < 0.01$ ) are sufficient to be confident in test-retest reliability.

#### 5.4.2.2 Test retest recap

Here are some summary notions for retest reliability:

- increases in the interval will lower the reliability coefficient,
- an experimental intervention that is designed to impact the retest assessment will lower the reliability coefficient,
- state measures will have lower retest coefficients than trait measures,
- and, the three phenomena above all interact with each other

Revelle and Condon's [2019a; 2019b] materials elaborate on this further. Their Table 1 is especially helpful. In addition to the myriad of vignettes used to illustrate issues with state, trait, items, whole scale, and so forth, there are demonstrations for duplicated items, assessing for consistency, and parallel/alternate forms.

If you are asking, "Hey, is parallel/alternate forms really a variant of test retest?" Great question! In fact, split-half could be seen as test-retest! Once you get in the weeds, the distinctions become less clear.

### 5.4.3 Interrater Reliability

#### 5.4.3.1 Cohen's kappa

Cohen's kappa coefficient is used to calculate proportions of agreement corrected for chance. This type of analysis occurs in research designs where there is some kind of (usually) categorical designation of a response. I don't have an outside research vignette for this. In the past, I was involved in research where members of the research team coded counselor utterances according to Hill's *helping skills* system designed by Clara Hill [Hill, 2020]. In the helping skills system, 15 different helping skills are divided into three larger groups that generally reflect the counseling trajectory: *exploration, insight, action*. One of our analyses coded counselor utterances into these three categories. Let's look at a fabricated (not based on any real data) simulation where four raters each evaluated 12 counselor utterances (that represent the arch of a nonsensically speedy counseling session).

```
Rater1 <- c("exploration", "exploration", "exploration", "exploration",
      "exploration", "exploration", "insight", "insight", "action", "action",
      "action", "action")
Rater2 <- c("exploration", "exploration", "exploration", "insight", "exploration",
      "insight", "exploration", "exploration", "exploration", "action", "exploration",
      "action")
Rater3 <- c("exploration", "insight", "exploration", "exploration", "exploration",
      "exploration", "exploration", "insight", "insight", "insight", "action",
      "action")
Rater4 <- c("exploration", "exploration", "exploration", "exploration",
      "exploration", "exploration", "exploration", "exploration", "exploration",
      "action", "action", "action")
ratings <- data.frame(Rater1, Rater2, Rater3, Rater4)
```

Historically, kappa could only be calculated for 2 raters at a time. Presently, though, it appears there can be any number of raters and the average agreement is reported.

Let's take a look at the data, then run the analysis, and interpret the results.

```
psych::cohen.kappa(ratings)
```

Cohen Kappa (below the diagonal) and Weighted Kappa (above the diagonal)  
For confidence intervals and detail print with all=TRUE

	Rater1	Rater2	Rater3	Rater4
Rater1	1.00	0.40	0.21	0.62
Rater2	0.14	1.00	0.00	0.57
Rater3	0.48	0.00	1.00	0.30
Rater4	0.54	0.45	0.43	1.00

Average Cohen kappa for all raters 0.34  
Average weighted kappa for all raters 0.35

Kappa can range from -1.00 to 1.00.

- $K = .00$  indicates that the observed agreement is exactly equal to the agreement that could be observed by chance.
- Negative kappa indicates that observed kappa is less than the expected chance agreement.
- $K = 1.00$  equals perfect agreement between judges.

There are commonly understood concerns about using kappa:

- Research teams typically set an expected standard (e.g., .85) and train raters until kappa is achieved.
  - In lengthy projects, rating agreement is rechecked periodically; if necessary there is retraining.
- Obtaining an acceptable kappa becomes difficult as the number of categories increases.
  - An example is Hill's *Helping Skills System* when all 15 categories; we chose to use the three categories (into which the 15 categories are subsumed).
- It is also difficult to obtain an adequate kappa when *infrequent* categories (e.g., "insight") exist.

Our kappa of .35 indicates that this rating team has a 35% chance of agreement, corrected for by chance. This is substantially below the standard. Let's imagine that the team spends time with their dictionaries, examines common errors, and makes some decision rules.

Here's the resimulation of "improved" agreement.

```
Rater1b <- c("exploration", "exploration", "exploration", "exploration",
           "exploration", "exploration", "insight", "insight", "insight", "action",
           "action", "action")
Rater2b <- c("exploration", "exploration", "exploration", "exploration",
           "exploration", "insight", "insight", "insight", "exploration", "action",
           "action", "action")
Rater3b <- c("exploration", "exploration", "exploration", "exploration",
           "exploration", "exploration", "exploration", "insight", "insight",
           "insight", "action", "action")
Rater4b <- c("exploration", "exploration", "exploration", "exploration",
           "exploration", "exploration", "exploration", "exploration", "insight",
           "action", "action", "action")
after_training <- data.frame(Rater1b, Rater2b, Rater3b, Rater4b)
```

Now run it again.

```
psych::cohen.kappa(after_training)
```

Warning in cohen.kappa1(x1, w = w, n.obs = n.obs, alpha = alpha, levels =
levels): upper or lower confidence interval exceed abs(1) and set to +/- 1.

Warning in cohen.kappa1(x1, w = w, n.obs = n.obs, alpha = alpha, levels = levels): upper or lower confidence interval exceed abs(1) and set to +/- 1.

Warning in cohen.kappa1(x1, w = w, n.obs = n.obs, alpha = alpha, levels = levels): upper or lower confidence interval exceed abs(1) and set to +/- 1.

Warning in cohen.kappa1(x1, w = w, n.obs = n.obs, alpha = alpha, levels = levels): upper or lower confidence interval exceed abs(1) and set to +/- 1.

Cohen Kappa (below the diagonal) and Weighted Kappa (above the diagonal)

For confidence intervals and detail print with all=TRUE

	Rater1b	Rater2b	Rater3b	Rater4b
Rater1b	1.00	0.83	0.55	0.80
Rater2b	0.73	1.00	0.36	0.60
Rater3b	0.72	0.45	1.00	0.46
Rater4b	0.71	0.43	0.70	1.00

Average Cohen kappa for all raters 0.62

Average weighted kappa for all raters 0.6

We observe improved scores, but this team needs more training if we aspire to a kappa of 0.85!

### 5.4.3.2 Intraclass correlation (ICC)

Another option for interrater reliability is the intraclass correlation (ICC). This is the same ICC we use in multilevel modeling! The ICC is used when we have numerical ratings.

In our fabricated vignette below, five raters are evaluating the campus climate for LGBTQIA+ individuals for 10 units/departments on a college campus. Using the ICC can help us determine the degree of leniency and variability within judges.

Below is a simulation of the data (you can ignore this)...

```
Rater1 <- c(1, 1, 1, 4, 2, 3, 1, 3, 3, 5)
Rater2 <- c(1, 1, 2, 1, 4, 4, 4, 4, 5, 5)
Rater3 <- c(3, 3, 3, 2, 3, 3, 6, 4, 4, 5)
Rater4 <- c(3, 5, 4, 2, 3, 6, 6, 6, 5, 5)
Rater5 <- c(2, 3, 3, 3, 4, 4, 4, 4, 5, 5)
ICC_df <- data.frame(Rater1, Rater2, Rater3, Rater4, Rater5)
```

```
#If the code below will not run remove the hashtags from the two lines of code below to install
```

```
#tools::package_dependencies("Matrix", which = "LinkingTo", reverse = TRUE)[[1L]]
#install.packages("lme4", type = "source")
```

We can use the *psych::ICC* function to obtain the ICC values.

```
# psych::ICC(ICC_df [1:10,1:5], lmer = TRUE) #find the ICCs for the
# 10 campus units and 5 judges
psych::ICC(ICC_df, missing = TRUE, alpha = 0.05, lmer = TRUE, check.keys = FALSE)
```

Call: psych::ICC(x = ICC\_df, missing = TRUE, alpha = 0.05, lmer = TRUE,  
check.keys = FALSE)

#### Intraclass correlation coefficients

	type	ICC	F	df1	df2	p	lower bound	upper bound
Single_raters_absolute	ICC1	0.34	3.5	9	40	0.00259	0.082	0.70
Single_random_raters	ICC2	0.37	5.4	9	36	0.00011	0.118	0.71
Single_fixed_raters	ICC3	0.47	5.4	9	36	0.00011	0.188	0.78
Average_raters_absolute	ICC1k	0.72	3.5	9	40	0.00259	0.308	0.92
Average_random_raters	ICC2k	0.74	5.4	9	36	0.00011	0.400	0.92
Average_fixed_raters	ICC3k	0.81	5.4	9	36	0.00011	0.537	0.95

Number of subjects = 10      Number of Judges = 5

See the help file for a discussion of the other 4 McGraw and Wong estimates,

In the output, reliability for a single judge  $ICC_1$  is the ratio of person variance to total variance. Reliability for multiple judges  $ICC_{1k}$  adjusts the residual variance by the number of judges.

The ICC function reports six reliability coefficients: 3 for the case of single judges and 3 for the case of multiple judges. It also reports the results in terms of a traditional ANOVA as well as a mixed effects linear model. Additionally, confidence intervals are reported.

Like most correlation coefficients, the ICC ranges from 0 to 1.

- An ICC close to 1 indicates high similarity between values from the same group.
- An ICC close to zero means that values from the same group are not similar.

## 5.5 What do we do with these coefficients?

### 5.5.1 Corrections for attenuation

Circa 1904, Spearman created the reliability coefficient out of a need to adjust observed correlations between related constructs for the error of measurement in each construct. This is only appropriate if the measure is seen as the expected value of a single underlying construct. However, “under the hood,” SEM programs model the pattern of observed correlations in terms of a measurement (reliability) model as well as a structural (validity) model.

### 5.5.2 Predicting true scores (and their CIs)

True scores remain unknown and so the reliability coefficient is used in a couple of ways to estimate the true score (and the confidence interval [CI] around that true score).

Take a quick look at the formula for predicting a true score and observe that the reliability coefficient is used within. It generally serves to nudge the observed score a bit closer to the mean:  $T' = (1 - r_{xx})\bar{X} + r_{xx}X$

The CI around that true score includes some estimate of standard error:  $CI_{95} = T' + / - z_{cv}(s_e)$ . Two estimates are commonly used. One is the standard error of estimate  $s_e = s_x \sqrt{r_{xx}(1 - r_{xx})}$  (i.e., the standard deviation of predicted true scores for a given observed score). Another is the standard error of measurement ( $s_m = s_x \sqrt{1 - r_{xx}}$ ) (i.e., an estimate of the amount of variation to be expected in test scores; aka, the standard deviation of the errors of measurement).

*I can hear you asking* What is the difference between  $s_e$  and  $s_m$ ?

- Because  $r_{xx}$  is almost always a fraction,  $s_e$  is smaller than  $s_m$ .
- When the reliability is high, the two standard errors are fairly similar to each other.
- Using  $s_m$  will result in wider confidence intervals.

### 5.5.3 How do I keep it all straight?

Table 1 in Revelle and Condon's [Revelle and Condon, 2019a] article helps us connect the type of reliability we are seeking with the statistic(s) and the R function within the *psych* package.

## 5.6 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The practice problems are the start of a larger project that spans multiple lessons. Therefore, if possible, please use a dataset that has item-level data for which there is a theorized total scale score as well as two or more subscales. With each of these options I encourage you to:

- Format (i.e., rescore if necessary) a dataset so that it is possible to calculate estimates of internal consistency
- Calculate and report the alpha coefficient for a total scale scores and subscales (if the scale has them)
- Calculate and report  $\omega_t$  and  $\omega_h$ . With these two determine what proportion of the variance is due to all the factors, error, and  $g$ .
- Calculate total and subscale scores.
- Describe other reliability estimates that would be appropriate for the measure you are evaluating.

### 5.6.1 Problem #1: Play around with this simulation.

If evaluating internal consistency is new to you, copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results. Perhaps you just change the number in "set.seed(210827)" from 210827 to something else. Your results should parallel those obtained in the lecture, making it easier for you to check your work as you go.

### 5.6.2 Problem #2: Use the data from the live ReCentering Psych Stats survey.

The script below pulls live data directly from the ReCentering Psych Stats survey on Qualtrics. As described in the [Scrubbing and Scoring chapters](#) of the ReCentering Psych Stats Multivariate Modeling volume, the Perceptions of the LGBTQ College Campus Climate Scale [Szymanski and Bissonette, 2020] was included (LGBTQ) and further adapted to assess perceptions of campus climate for Black students (BLst), non-Black students of color (nBSoC), international students (INTst), and students disabilities (wDIS). Consider conducting the analyses on one of these scales or merging them together.

```
library(tidyverse)
# only have to run this ONCE to draw from the same Qualtrics
# account...but will need to get different token if you are changing
# between accounts
library(qualtRics)
# qualtrics_api_credentials(api_key =
# 'mUgPMYsSYkiWpMFkwHale1QE5HNmh5LRUaA8d9PDg', base_url =
# 'spupsych.az1.qualtrics.com', overwrite = TRUE, install = TRUE)
QTRX_df <- qualtRics::fetch_survey(surveyID = "SV_b2cClqA1LGQ6nLU", time_zone = NULL,
verbose = FALSE, label = FALSE, convert = FALSE, force_request = TRUE,
import_id = FALSE)
climate_df <- QTRX_df %>%
  select("BLst_1", "BLst_2", "BLst_3", "BLst_4", "BLst_5", "BLst_6",
         "nBSoC_1", "nBSoC_2", "nBSoC_3", "nBSoC_4", "nBSoC_5", "nBSoC_6",
         "INTst_1", "INTst_2", "INTst_3", "INTst_4", "INTst_5", "INTst_6",
         "wDIS_1", "wDIS_2", "wDIS_3", "wDIS_4", "wDIS_5", "wDIS_6", "LGBTQ_1",
         "LGBTQ_2", "LGBTQ_3", "LGBTQ_4", "LGBTQ_5", "LGBTQ_6")
# Item numbers are supported with the following items: _1 'My campus
# unit provides a supportive environment for ___ students' _2
# '_____ is visible in my campus unit' _3 'Negative attitudes
# toward persons who are ___ are openly expressed in my campus
# unit.' _4 'My campus unit is unresponsive to the needs of ___
# students.' _5 'Students who are_____ are harassed in my campus
# unit.' _6 'My campus unit is cold and uncaring toward ___
# students.'

# Item 1 on each subscale should be reverse coded. The College
# Response scale is composed of items 1, 4, 6, The Stigma scale is
# composed of items 2,3, 5
```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do).

```
# write the simulated data as a .csv write.table(climate_df,
# file='climate_df.csv', sep=',', col.names=TRUE, row.names=FALSE)
# bring back the simulated dat from a .csv file climate_df <-
# read.csv ('climate_df.csv', header = TRUE)
```

```
# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(climate_df, 'climate_df.rds') bring back the simulated
# dat from an .rds file climate_df <- readRDS('climate_df.rds')
```

### 5.6.3 Problem #3: Try something entirely new.

Complete the same steps using data for which you have permission and access. This might be data of your own, from your lab, simulated from an article, or located on an open repository.

### 5.6.4 Grading Rubric

Assignment Component	Points Possible	Points Earned
1. Check and, if needed, format and score data	5	_____
2. Calculate and report the alpha coefficient for a total scale scores and subscales (if the scale has them)	5	_____
3. Calculate and report $\omega_t$ and $\omega_h$ . With these two determine what proportion of the variance is due to all the factors, error, and $g$ .	5	_____
4. Calculate total and subscale scores.	5	_____
5. Describe other reliability estimates that would be appropriate for the measure you are evaluating.	5	_____
6. Explanation to grader	5	_____
<b>Totals</b>	<b>30</b>	_____

## 5.7 Homeworked Example

### Screencast Link

For more information about the data used in this homeworked example, please refer to the description and codebook located at the end of the [introduction](#) in first volume of ReCentering Psych Stats.

As a brief review, this data is part of an IRB-approved study, with consent to use in teaching demonstrations and to be made available to the general public via the open science framework. Hence, it is appropriate to use in this context. You will notice there are student- and teacher- IDs. These numbers are not actual student and teacher IDs, rather they were further re-identified so that they could not be connected to actual people.

Because this is an actual dataset, if you wish to work the problem along with me, you will need to download the [ReC.rds](#) data file from the Worked\_Examples folder in the ReC\_Psychometrics project on the GitHub.

The course evaluation items can be divided into three subscales:

- **Valued by the student** includes the items: ValObjectives, IncrUnderstanding, IncrInterest
- **Traditional pedagogy** includes the items: ClearResponsibilities, EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation
- **Socially responsive pedagogy** includes the items: InclusvClassrm, EquitableEval, MultPerspectives, DEIIntegration

In this homework focused on reliability we will report alpha coefficients for total scale score and subscale scores. We'll also calculate omega total and omega hierarchical and determine what proportion of variance is due to all the factors, error, and  $g$ . Finally, we'll calculate total and subscale scores.

### 5.7.1 Check and, if needed, format the data

```
big <- readRDS("ReC.rds")
```

Let's check the structure...

```
str(big)
```

```
Classes 'data.table' and 'data.frame': 310 obs. of 33 variables:
 $ deID                  : int 1 2 3 4 5 6 7 8 9 10 ...
 $ CourseID               : int 57085635 57085635 57085635 57085635 57085635 57085635 57085635 57085635 ...
 $ Dept                   : chr "CPY" "CPY" "CPY" "CPY" ...
 $ Course                 : Factor w/ 3 levels "Psychometrics",...: 2 2 2 2 2 2 2 2 2 2 ...
 $ StatsPkg                : Factor w/ 2 levels "SPSS","R": 2 2 2 2 2 2 2 2 2 2 ...
 $ Centering               : Factor w/ 2 levels "Pre","Re": 2 2 2 2 2 2 2 2 2 2 ...
 $ Year                    : int 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ Quarter                 : chr "Fall" "Fall" "Fall" "Fall" ...
 $ IncrInterest             : int 5 3 4 2 4 3 5 3 2 5 ...
 $ IncrUnderstanding         : int 2 3 4 3 4 4 5 2 4 5 ...
 $ ValObjectives            : int 5 5 4 4 5 5 5 5 4 5 ...
 $ ApprAssignments           : int 5 4 4 4 5 3 5 3 3 5 ...
 $ EffectiveAnswers          : int 5 3 5 3 5 3 4 3 2 3 ...
 $ Respectful                : int 5 5 4 5 5 4 5 4 5 5 ...
 $ ClearResponsibilities     : int 5 5 4 4 5 4 5 4 4 5 ...
 $ Feedback                  : int 5 3 4 2 5 NA 5 4 4 5 ...
 $ OvInstructor              : int 5 4 4 3 5 3 5 4 3 5 ...
 $ MultPerspectives           : int 5 5 4 5 5 4 5 5 5 5 ...
 $ OvCourse                  : int 3 4 4 3 5 3 5 3 2 5 ...
 $ InclusvClassrm            : int 5 5 5 5 5 4 5 5 5 4 5 ...
 $ DEIIntegration             : int 5 5 5 5 5 4 5 5 5 5 5 ...
 $ ClearPresentation           : int 4 4 4 2 5 3 4 4 4 5 ...
 $ ApprWorkload               : int 5 5 3 4 4 2 5 4 4 5 ...
 $ MyContribution              : int 4 4 4 4 5 4 4 3 4 5 ...
 $ InspiredInterest            : int 5 3 4 3 5 3 5 4 4 5 ...
```

```
$ Faith : int 5 NA 4 2 NA NA 4 4 4 NA ...
$ EquitableEval : int 5 5 3 5 5 3 5 5 3 5 ...
$ SPFC.Decolonize.Opt.Out: chr "" "" "" ...
$ ProgramYear : Factor w/ 3 levels "Second","Transition",...: 3 3 3 3 3 3 3 3 3 ...
$ ClearOrganization : int 3 4 3 4 4 4 5 4 4 5 ...
$ RegPrepare : int 5 4 4 4 4 3 4 4 4 5 ...
$ EffectiveLearning : int 2 4 3 4 4 2 5 3 2 5 ...
$ AccessibleInstructor : int 5 4 4 4 5 4 5 4 5 5 ...
- attr(*, ".internal.selfref")=<externalptr>
```

Let's create a df with the items only.

```
library(tidyverse)
items <- big %>%
  dplyr::select(ValObjectives, IncrUnderstanding, IncrInterest, ClearResponsibilities,
    EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation,
    MultPerspectives, InclusvClassrm, DEIintegration, EquitableEval)
```

### 5.7.2 Calculate and report the alpha coefficient for a total scale score and subscales (if the scale has them)

```
psych::alpha(items)
```

```
Reliability analysis
Call: psych::alpha(x = items)
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.92	0.92	0.93	0.49	11	0.0065	4.3	0.61	0.48

```
95% confidence boundaries
      lower alpha upper
Feldt     0.90  0.92  0.93
Duhachek  0.91  0.92  0.93
```

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
ValObjectives	0.92	0.92	0.93	0.51	11.3	0.0067	0.016	
IncrUnderstanding	0.91	0.91	0.92	0.49	10.6	0.0070	0.016	
IncrInterest	0.91	0.91	0.92	0.49	10.4	0.0070	0.018	
ClearResponsibilities	0.91	0.91	0.92	0.48	10.0	0.0073	0.015	
EffectiveAnswers	0.91	0.91	0.92	0.48	10.0	0.0074	0.016	
Feedback	0.91	0.91	0.92	0.48	10.3	0.0071	0.018	
ClearOrganization	0.91	0.91	0.92	0.48	10.2	0.0073	0.016	
ClearPresentation	0.91	0.91	0.92	0.47	9.7	0.0076	0.015	

MultPerspectives	0.91	0.91	0.92	0.48	10.0	0.0073	0.017
InclusvClassrm	0.91	0.91	0.92	0.49	10.6	0.0069	0.018
DEIIntegration	0.92	0.92	0.93	0.52	11.8	0.0063	0.011
EquitableEval	0.91	0.91	0.93	0.49	10.5	0.0070	0.018
med.r							
ValObjectives	0.53						
IncrUnderstanding	0.50						
IncrInterest	0.48						
ClearResponsibilities	0.48						
EffectiveAnswers	0.48						
Feedback	0.48						
ClearOrganization	0.48						
ClearPresentation	0.47						
MultPerspectives	0.47						
InclusvClassrm	0.52						
DEIIntegration	0.53						
EquitableEval	0.48						

## Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
ValObjectives	309	0.59	0.61	0.55	0.53	4.5	0.61
IncrUnderstanding	309	0.71	0.70	0.67	0.64	4.3	0.82
IncrInterest	308	0.75	0.73	0.71	0.68	3.9	0.99
ClearResponsibilities	307	0.80	0.80	0.79	0.75	4.4	0.82
EffectiveAnswers	308	0.80	0.79	0.78	0.75	4.4	0.83
Feedback	304	0.75	0.75	0.72	0.69	4.2	0.88
ClearOrganization	309	0.79	0.77	0.75	0.72	4.0	1.08
ClearPresentation	309	0.85	0.84	0.83	0.80	4.2	0.92
MultPerspectives	305	0.79	0.80	0.78	0.75	4.4	0.84
InclusvClassrm	301	0.68	0.70	0.67	0.62	4.6	0.68
DEIIntegration	273	0.51	0.53	0.49	0.42	4.5	0.74
EquitableEval	308	0.70	0.72	0.69	0.66	4.6	0.63

## Non missing response frequency for each item

	1	2	3	4	5	miss
ValObjectives	0.00	0.01	0.03	0.39	0.57	0.00
IncrUnderstanding	0.01	0.04	0.07	0.44	0.45	0.00
IncrInterest	0.02	0.09	0.14	0.44	0.31	0.01
ClearResponsibilities	0.01	0.02	0.07	0.31	0.59	0.01
EffectiveAnswers	0.01	0.02	0.08	0.36	0.53	0.01
Feedback	0.01	0.05	0.10	0.39	0.46	0.02
ClearOrganization	0.04	0.07	0.10	0.41	0.38	0.00
ClearPresentation	0.02	0.05	0.07	0.40	0.46	0.00
MultPerspectives	0.02	0.02	0.08	0.33	0.56	0.02
InclusvClassrm	0.01	0.01	0.05	0.23	0.70	0.03
DEIIntegration	0.00	0.01	0.10	0.22	0.67	0.12
EquitableEval	0.00	0.01	0.03	0.32	0.63	0.01

Total scale score alpha is 0.92

### 5.7.3 Subscale alphas

In the lecture, I created baby dfs of the subscales and ran the alpha on those; another option is to use concatenated lists of variables (i.e., variable vectors). Later, we can also use these to score the subscales.

```
ValuedVars <- c("ValObjectives", "IncrUnderstanding", "IncrInterest")
TradPedVars <- c("ClearResponsibilities", "EffectiveAnswers", "Feedback",
                 "ClearOrganization", "ClearPresentation")
SRPedVars <- c("InclusvClassrm", "EquitableEval", "MultPerspectives", "DEIintegration")

psych::alpha(items[, ValuedVars])
```

```
Reliability analysis
Call: psych::alpha(x = items[, ValuedVars])

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean    sd median_r
 0.77      0.77      0.71      0.53 3.4 0.02  4.2 0.68      0.48

  95% confidence boundaries
        lower alpha upper
Feldt     0.72  0.77  0.81
Duhachek 0.73  0.77  0.81

Reliability if an item is dropped:
  raw_alpha std.alpha G6(smc) average_r S/N   alpha se var.r
ValObjectives       0.80      0.81      0.68      0.68 4.3  0.022  NA
IncrUnderstanding    0.60      0.65      0.48      0.48 1.8  0.040  NA
IncrInterest        0.59      0.61      0.44      0.44 1.6  0.044  NA
  med.r
ValObjectives       0.68
IncrUnderstanding    0.48
IncrInterest         0.44

Item statistics
  n raw.r std.r r.cor r.drop mean    sd
ValObjectives     309  0.71  0.77  0.55   0.50  4.5 0.61
IncrUnderstanding 309  0.86  0.85  0.76   0.68  4.3 0.82
IncrInterest      308  0.90  0.87  0.79   0.70  3.9 0.99

Non missing response frequency for each item
  1   2   3   4   5 miss
ValObjectives     0.00 0.01 0.03 0.39 0.57 0.00
```

```
IncrUnderstanding 0.01 0.04 0.07 0.44 0.45 0.00
IncrInterest      0.02 0.09 0.14 0.44 0.31 0.01
```

Alpha for the Valued-by-Me dimension is .77

```
psych::alpha(items[, TradPedVars])
```

```
Reliability analysis
Call: psych::alpha(x = items[, TradPedVars])
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.89	0.9	0.88	0.64	8.8	0.0094	4.3	0.76	0.65

	95% confidence boundaries		
	lower	alpha	upper
Feldt	0.87	0.89	0.91
Duhachek	0.88	0.89	0.91

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
ClearResponsibilities	0.86	0.86	0.84	0.62	6.4	0.013	0.0054		0.59
EffectiveAnswers	0.87	0.87	0.84	0.63	6.8	0.012	0.0045		0.65
Feedback	0.89	0.89	0.87	0.68	8.4	0.010	0.0016		0.69
ClearOrganization	0.88	0.88	0.85	0.64	7.2	0.012	0.0044		0.66
ClearPresentation	0.86	0.87	0.83	0.62	6.5	0.013	0.0030		0.62

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
ClearResponsibilities	307	0.87	0.87	0.84	0.79	4.4	0.82
EffectiveAnswers	308	0.84	0.85	0.81	0.76	4.4	0.83
Feedback	304	0.78	0.79	0.70	0.66	4.2	0.88
ClearOrganization	309	0.85	0.83	0.78	0.74	4.0	1.08
ClearPresentation	309	0.87	0.87	0.83	0.78	4.2	0.92

Non missing response frequency for each item

	1	2	3	4	5	miss
ClearResponsibilities	0.01	0.02	0.07	0.31	0.59	0.01
EffectiveAnswers	0.01	0.02	0.08	0.36	0.53	0.01
Feedback	0.01	0.05	0.10	0.39	0.46	0.02
ClearOrganization	0.04	0.07	0.10	0.41	0.38	0.00
ClearPresentation	0.02	0.05	0.07	0.40	0.46	0.00

Alpha for Traditional Pedagogy dimension is .90

```
psych::alpha(items[, SRPedVars])
```

Reliability analysis

```
Call: psych::alpha(x = items[, SRPedVars])
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.81	0.81	0.78	0.52	4.3	0.017	4.5	0.58	0.54

95% confidence boundaries

	lower	alpha	upper
--	-------	-------	-------

Feldt	0.77	0.81	0.84
-------	------	------	------

Duhachek	0.77	0.81	0.84
----------	------	------	------

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
InclusvClassrm	0.74	0.74	0.67	0.49	2.9	0.025	0.0120	
EquitableEval	0.78	0.79	0.73	0.56	3.9	0.021	0.0034	
MultPerspectives	0.73	0.74	0.67	0.49	2.8	0.026	0.0153	
DEIIntegration	0.78	0.78	0.71	0.54	3.6	0.021	0.0044	
	med.r							
InclusvClassrm	0.50							
EquitableEval	0.57							
MultPerspectives	0.47							
DEIIntegration	0.57							

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
InclusvClassrm	301	0.82	0.83	0.76	0.67	4.6	0.68
EquitableEval	308	0.75	0.76	0.64	0.58	4.6	0.63
MultPerspectives	305	0.85	0.83	0.76	0.68	4.4	0.84
DEIIntegration	273	0.78	0.78	0.67	0.59	4.5	0.74

Non missing response frequency for each item

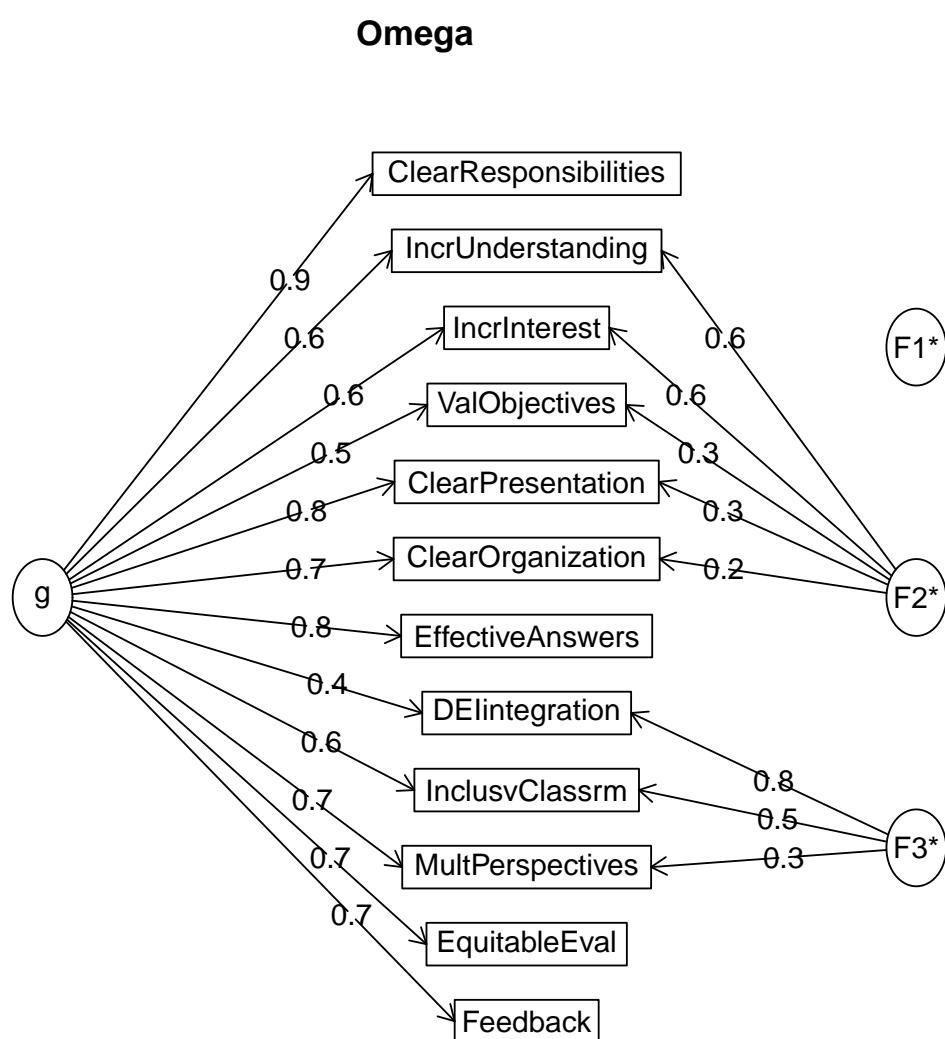
	1	2	3	4	5	miss
InclusvClassrm	0.01	0.01	0.05	0.23	0.70	0.03
EquitableEval	0.00	0.01	0.03	0.32	0.63	0.01
MultPerspectives	0.02	0.02	0.08	0.33	0.56	0.02
DEIIntegration	0.00	0.01	0.10	0.22	0.67	0.12

Alpha for the SCR Pedagogy dimension is .81

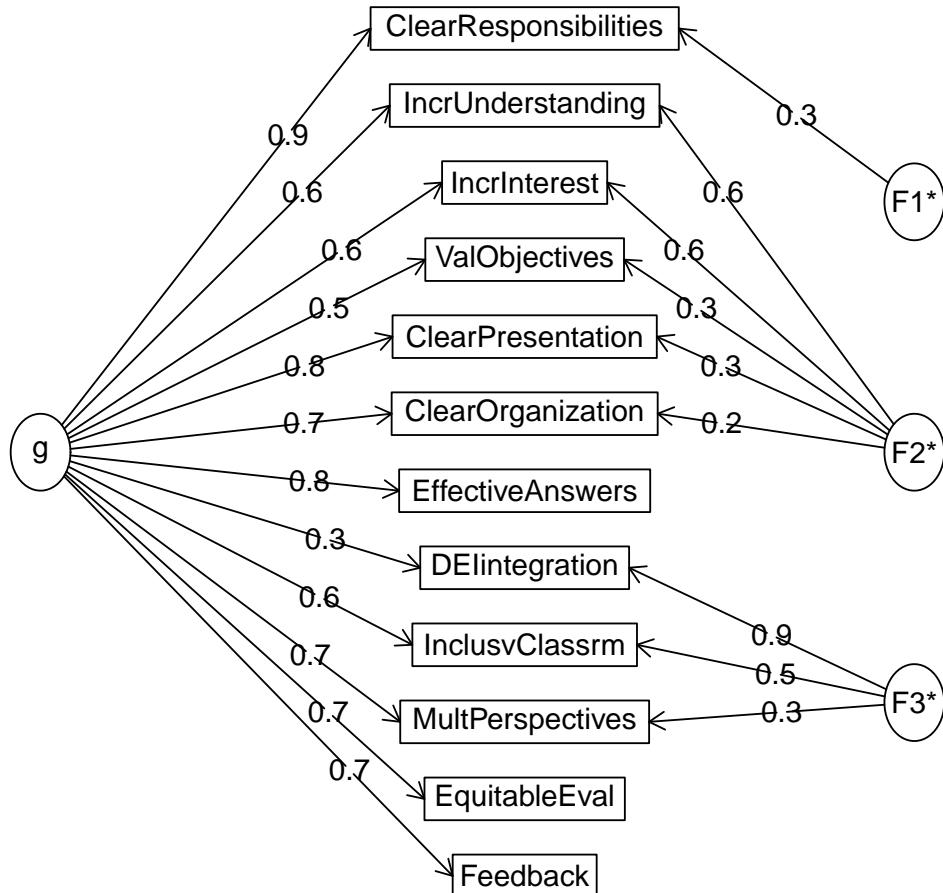
### 5.7.4 Calculate and report t and h

```
psych::omegaSem(items, nfactors = 3)
```

Warning in lav\_model\_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WARNING:  
 Could not compute standard errors! The information matrix could  
 not be inverted. This may be a symptom that the model is not  
 identified.



### Omega from SEM



```

Call: psych::omegaSem(m = items, nfactors = 3)
Omega
Call: omegah(m = m, nfactors = nfactors, fm = fm, key = key, flip = flip,
  digits = digits, title = title, sl = sl, labels = labels,
  plot = plot, n.obs = n.obs, rotate = rotate, Phi = Phi, option = option,
  covar = covar)
Alpha:          0.92
G.6:           0.93
Omega Hierarchical: 0.83
Omega H asymptotic: 0.88
Omega Total      0.94

```

	Schmid Leiman Factor loadings greater than 0.2						
	<i>g</i>	$F_{1*}$	$F_{2*}$	$F_{3*}$	<i>h</i> <sup>2</sup>	<i>u</i> <sup>2</sup>	<i>p</i> <sup>2</sup>
ValObjectives	0.47		0.32		0.33	0.67	0.67
IncrUnderstanding	0.57		0.60		0.69	0.31	0.47

IncrInterest	0.58	0.57	0.67	0.33	0.49
ClearResponsibilities	0.87		0.78	0.22	0.98
EffectiveAnswers	0.79		0.65	0.35	0.97
Feedback	0.73		0.56	0.44	0.96
ClearOrganization	0.75	0.23	0.62	0.38	0.90
ClearPresentation	0.81	0.30	0.74	0.26	0.88
MultPerspectives	0.75		0.30	0.65	0.35
InclusvClassrm	0.56		0.48	0.57	0.43
DEIIntegration	0.37		0.82	0.80	0.20
EquitableEval	0.70		0.52	0.48	0.94

With Sums of squares of:

	$g$	$F1^*$	$F2^*$	$F3^*$
5.51	0.04	0.98	1.06	

general/max 5.19 max/min = 26.42  
mean percent general = 0.74 with sd = 0.26 and cv of 0.36  
Explained Common Variance of the general factor = 0.73

The degrees of freedom are 33 and the fit is 0.25

The number of observations was 310 with Chi Square = 76.65 with prob < 0.000025

The root mean square of the residuals is 0.02

The df corrected root mean square of the residuals is 0.03

RMSEA index = 0.065 and the 10 % confidence intervals are 0.046 0.085

BIC = -112.66

Compare this with the adequacy of just a general factor and no group factors

The degrees of freedom for just the general factor are 54 and the fit is 1.37

The number of observations was 310 with Chi Square = 415 with prob < 0.000000000000000000000000000000

The root mean square of the residuals is 0.1

The df corrected root mean square of the residuals is 0.11

RMSEA index = 0.147 and the 10 % confidence intervals are 0.134 0.16

BIC = 105.22

## Measures of factor score adequacy

g F1\* F2\* F3\*

Correlation of scores with factors                    0.95    0.13    0.81    0.89

Multiple R square of scores with factors 0.91 0.02 0.66 0.79

Minimum correlation of factor score estimates 0.82 -0.96 0.32 0.58

Total, General and Subset omega for each subset

g F1\* F2\* F3\*

Omega total for total scores and subscales 0.94 0.77 0.90 0.87

Omega general for total scores and subscales 0.83 0.76 0.69 0.64

Omega group for total scores and subscales 0.11 0.01 0.20 0.23

The following analyses were done using the `lavaan` package

Omega Hierarchical from a confirmatory model using sem = 0.82

Omega Total from a confirmatory model using sem = 0.94

With loadings of

	g	F1*	F2*	F3*	h2	u2	p2
ValObjectives	0.47		0.32		0.33	0.67	0.67
IncrUnderstanding	0.56		0.62		0.70	0.30	0.45
IncrInterest	0.57		0.56		0.64	0.36	0.51
ClearResponsibilities	0.86	0.32			0.84	0.16	0.88
EffectiveAnswers	0.80				0.65	0.35	0.98
Feedback	0.73				0.55	0.45	0.97
ClearOrganization	0.75		0.22		0.61	0.39	0.92
ClearPresentation	0.82		0.27		0.74	0.26	0.91
MultPerspectives	0.74			0.30	0.64	0.36	0.86
InclusvClassrm	0.56			0.49	0.56	0.44	0.56
DEIIntegration	0.33			0.87	0.87	0.13	0.13
EquitableEval	0.69				0.51	0.49	0.93

With sum of squared loadings of:

g	F1*	F2*	F3*
5.46	0.10	0.94	1.14

The degrees of freedom of the confirmatory model are 42 and the fit is 110.6184 with p = 0  
general/max 4.8 max/min = 10.87

mean percent general = 0.73 with sd = 0.27 and cv of 0.37

Explained Common Variance of the general factor = 0.71

Measures of factor score adequacy

	g	F1*	F2*	F3*
Correlation of scores with factors	0.95	0.57	0.82	0.95
Multiple R square of scores with factors	0.91	0.32	0.67	0.90
Minimum correlation of factor score estimates	0.82	-0.36	0.35	0.80

Total, General and Subset omega for each subset

	g	F1*	F2*	F3*
Omega total for total scores and subscales	0.94	0.84	0.89	0.87
Omega general for total scores and subscales	0.82	0.74	0.70	0.62
Omega group for total scores and subscales	0.11	0.10	0.20	0.26

To get the standard sem fit statistics, ask for summary on the fitted object

I'm reporting the values below the, "The following analyses were done using the lavaan package":

Omega total = .94 (omega total values > .80 are an indicator of good reliability). Interpretation: 94% of the variance in the total scale is due to the factors and the balance (6%) is due to error.

Omega hierarchical estimates the proportion of variance in the overall course evaluation score attributable to the general factors (thus treating the subscales as error). Omega h for the overall course evaluation score was .82

**5.7.5 With these two determine what proportion of the variance is due to all the factors, error, and g.**

A quick calculation with omega h (.82) and omega total (.94)

```
.82/.94
```

```
[1] 0.8723404
```

let's us know that 87% of the reliable variance in the overall course evaluation score is attributable to the general factor

**5.7.6 Calculate total and subscale scores.**

This code uses the variable vectors I created above.

```
items$Valued <- sjstats::mean_n(items[, ValuedVars], 0.75)
items$TradPed <- sjstats::mean_n(items[, TradPedVars], 0.75)
items$SCRPed <- sjstats::mean_n(items[, SCRPedVars], 0.75)
items$Total <- sjstats::mean_n(items, 0.75)
```

```
scores <- items %>%
  dplyr::select(Valued, TradPed, SCRPed, Total)

psych::describe(scores)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
Valued	1	309	4.25	0.68	4.33	4.32	0.50	1.67	5	3.33	-0.90	0.57
TradPed	2	307	4.25	0.76	4.40	4.37	0.59	1.00	5	4.00	-1.42	2.48
SCRPed	3	299	4.52	0.58	4.75	4.61	0.37	2.25	5	2.75	-1.25	1.33
Total	4	308	4.34	0.60	4.41	4.41	0.62	1.83	5	3.17	-1.07	1.12
	se											
Valued		0.04										
TradPed		0.04										
SCRPed		0.03										
Total		0.03										

**5.7.7 Describe other reliability estimates that would be appropriate for the measure you are evaluating.**

These scales are for the purposes of course evaluations. In their development, it might be helpful to give it at the end of a single course and then again a few weeks later to determine test-retest reliability.



# Chapter 6

# Item Analysis for Educational Achievement Tests (Exams)

## [Screencasted Lecture Link](#)

In this lecture I walk through some procedures for analyzing the quality of multiple choice (including true/false) exam items. We look at item difficulty and item discrimination. We also look at item coverage as it relates to the learning objectives for an educational endeavor.

## 6.1 Navigating this Lesson

There is about one hour of lecture. If you work through the materials with me it would be plan for an additional 30 minutes.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

### 6.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Provide a rationale for why having a *test bank* might be a good idea.
- Describe the effects of skewness on the interpretation of exam results.
- Evaluate the the quality of a multiple choice item on the basis of item difficulty, correlation, and discrimination.
- Discuss the challenges of identifying an *ideal* difficulty level for test items. Further elaborate how guessing, speeded tests, interitem correlations, and the purposes of the test influence the *ideal difficulty*.

### 6.1.2 Planning for Practice

Practice suggestions for this lesson encourage you to think about the exams in your life: those you might be taking; those you might be writing or proctoring.

### 6.1.3 Readings & Resources

Classic psychometric texts tend to not cover item analysis for achievement tests and/or they skip over these fundamentals and move straight to item response theory/Rasch modeling (IRT). After scouring the internet, I landed on these two resources as concise, accessible, summaries.

- Understanding item analysis. Office of Educational Assessment, University of Washington. Retrieved September 20, 2019. Retrieved from <https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/>
  - It is common for excellent instructions/descriptions to accompany the scoring software used by institutions. UW appears to use ScorePak and this resource provides both conceptual and interpretive information.
- Revelle, W. (2017). An overview of the psych package. Retrieved from <http://personality-project.org/r/overview.pdf>
  - Pages 85-85 provide a vignette for conducting item analysis on multiple choice items.

### 6.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(psych)){install.packages('psych')}
```

## 6.2 Research Vignette

This lesson's research vignette is from my own class. Especially in the early years of my teaching, I gave high(er) stakes mid-term and final exams. There were usually 40 (or so) multiple choice or true/false items, 2-3 applied problems or short essays, and 1 longer essay. Today's vignette are an array of exam items from a statistics exam that demonstrate the desirable and undesirable elements we want in objective items.

## 6.3 Item Analysis in the Educational/Achievement Context

Multiple choice, true/false, and other *objectively* formatted/scored items are part-n-parcel to educational/achievement assessment. But how do we know if the items are performing the way they

should? This lecture focuses on item analysis in the context of multiple choice and true/false items. Using these practices can help you identify what selection of items you'd like for your exams. These can be critical tools in helping you improve your ability to assess student performance. In-so-doing, we walk through a bit of "what we used to do," to current common practices, to a glimpse of our future. We owe much of this to rapid advances in technology.

*Test banks* are instructor-created resources for developing/storing/protecting items for use in future exams. We create test banks when we carefully distribute/collect/protect items "that work" (from statistical perspective). Why would we want to do this?

- Once a test is "out" it's out. Instructors can presume that resourceful students are using it to study; yet all students won't have equal access to it.
- Developing "good" items takes a good deal of time; does the instructor want to redo this each term?
- Should we be piloting *all new items* on students each term and then having the debates about whether the item should be rescored?
  - Better is to introduce a proportion of new items each year and evaluate them for inclusion in the test bank; EPPP, SAT, GRE do this.
- A challenge is providing students appropriate study tools – old exams are favorites of students (but maybe there are other ways – worksheets, Jeopardy).

The conceptual portions of this lecture, particularly the interpretation of the difficulty and discrimination statistics are based in Anastasi's work [[Anastasi and Urbina, 1997](#)]

### 6.3.1 And now a quiz! Please take it.

Let's start with some items from an early version of the exam I gave when I taught CPY7020/Statistical Methods.

**Item 5** A grouping variable such as men or women that uses dummy coding of 1 and 0 to categorize the groups is an example of \_\_\_\_\_ scaling.

- a) Nominal
- b) Ordinal
- c) Interval
- d) Ratio

**Item 11** The term "grade inflation" has frequently been applied to describe the distribution of grades in graduate school. Which of the following best describes this distribution.

- a) negatively skewed
- b) uniform/rectangular
- c) positively skewed and leptokurtic

- d) uniform and platykurtic

**Item 19** All distributions of Z-scores will have the identical

- a) Mean
- b) Variance
- c) Standard deviation
- d) All of the above

**Item 21** The most appropriate score for comparing scores across two or more distributions (e.g., exam scores in math and art classes) is the:

- a) mean
- b) percentile rank
- c) raw score
- d) z-score

**Item 37** Of the following, what statement best describes  $r^2 = .49$

- a) strong positive correlation
- b) strong positive or negative correlation
- c) weak positive or negative correlation
- d) weak negative correlation

**Item 38** When there are no ties among ranks, what is the relationship between the Spearman rho ( $\rho$ ) and the Pearson ( $r$ )?

- a)  $\rho = r$
- ii)  $\rho > r$
- a)  $\rho < r$
- b) no relationship

## 6.4 Item Difficulty

### 6.4.1 Percent passing

**Item difficulty index** is the proportion of test takers who answer an item correctly. It is calculated by dividing the number of people who passed the item (e.g., 55) by the total number of people (e.g., 100).

- If 55% pass an item, we write  $p = .55$
- The easier the item, the larger the percentage will be.

What is an ideal pass rate (and this “ideal” is the *statistical ideal* mostly for norm-referenced tests like the ACT, SAT, GRE)?

- The closer the difficulty of an item approaches 1.00 or 0, the less differential information about test takers it contributes.
  - If, out of 100 people, 50 pass an item and 50 fail ( $p = .50$ )...we have 50 X 50 or 2,500 paired comparisons or differential bits of information.
- How much information would we have for an item passed by:
  - 70% of the people ( $70 * 30 = ???$ )
  - 90% of the people ( $90 * 10 = ???$ )
- For maximum differentiation, one would choose all items at the .50 level (but hold up...)

### 6.4.2 Several factors prevent .50 from being the ideal difficulty level

**Speeded tests** complicate the interpretation of item difficulty because items are usually of equivalent difficulty and there are so many that no one could complete them all. Thus later items should be considered to be more difficult – but item difficulty is probably not the best assessment of item/scale quality.

**Guessing** the correct answer in true/false and multiple choice contexts interferes with the goal of  $p = .50$ . In a 1952 issue of *Psychometrika*, Lord provided this guide for optimal  $p$  values based on the number of choices in the objective context:

Optimal  $p$  values

Number of Choices	Optimal Mean Difficulty Level
2 (T/F)	0.85
3	0.77
4	0.74
5	0.69
Constructed response essay	0.5

The purpose of the testing changes the ideal difficulty level.

- If the test is *norm-referenced* (ACT, SAT, GRE), .50 is very useful.
- If the test is mastery oriented,  $p$  values may be as high as 0.90 since student performance is a function of repeated attempts with feedback.

**Item intercorrelations** impacts interpretation of item difficulty.

- The more homogeneous the test, the higher these intercorrelations will be. If all items were perfectly intercorrelated and all were of the .50 difficulty level:
  - the same 50 persons out of 100 would pass each item, that is,
  - half of the test takers would obtain perfect scores, the other half zero scores
- It is best to select items with a moderate spread of difficulty but whose AVERAGE difficulty level is .50
- The percentage of persons passing an item expresses the item difficulty in terms of which statistical scale of measurement? Is it nominal, ordinal, interval, or ratio?
  - Because of this issue, we can correctly indicate the rank order or relative difficulty of the items
  - However, we cannot infer that the difference in difficulty between Items 1 and 2 is equal to the difference between Items 2 and 3.
- We can make an *equal-interval inference* with the table of normal curve frequencies (i.e., translating the proportion to z-scores). Z-scores would be used as the units if an equal interval inference was required in the analysis. For example,
  - $p = .84$  is equal to  $-1\text{ SD}$
  - $p = .16$  is equal to  $+1\text{ SD}$

*Seem a little upside down? Recall that we are calculating the percent passing and starting the count “from the top.” So a relatively easy item where 84% passed, would have an standard deviation of -1.*

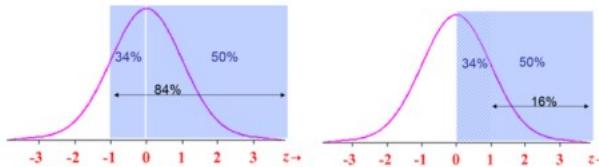


Figure 6.1: Image of graphs where  $p = .84$  and  $p = .16$

## 6.5 Item Discrimination

The degree to which an item differentiates correctly among test takers in the behavior that the test is designed to measure.

- the *criterion* can be internal or external to the test itself

- under some conditions, the two approaches lead to opposite results because (a) items chosen to maximize the validity of the test tend to be the ones rejected on the basis of internal consistency, and (b) rejecting items with low correlations with the total score tends to homogenize the test (we are more likely to keep items with the highest average intercorrelations)
- *internal* criteria maximizes internal consistency or homogeneity of the test.
  - Example: achievement test, where criteria is total score itself
- *external* criteria maximize the validity of an external criterion.
  - Example: a different assessment of the same ability being assessed

### 6.5.1 Index of Discrimination

- Compare the proportion of cases that pass an item in contrasting criterion groups
  - upper (U) and lower (L) criterion groups are selected from the extremes of the distribution
  - traditionally these groups are created from the 27% from each of those sides of the distribution
- This *index of discrimination (D)* can be expressed as a difference of raw frequencies (U - L), or (more conventionally) as the difference of percentages of those who scored it correctly in the upper 27% and lower 27% groups
  - when all members of the U group and none of the members of the L group pass, D = 100
  - when all members of the L group and none of the members of the U group pass, D = 0
  - optimum point at which these two conditions reach balance is with the upper and lower 27%

---

#### Optimal Discrimination

---

Difficulty	Discrimination
0.40 and larger	Excellent
0.30 - 0.39	Good
0.11 - 0.29	Fair
0.00 -0.10	Poor
Negative values	Mis-keyed or other major flaw

### 6.5.2 Application of Item Difficulty and Discrimination

Earlier I asked you to “take the quiz.” To keep it engaging, I encourage you to look at your own answers and compare them to “what happened” from in this actual exam administration. I will

KEY		A	A	D	D	B	A		
	ID	Missed	Score	Item5	Item11	Item19	Item21	Item37	Item38
	1	3	38	A	A	D	B	B	A
	10	3	38	A	C	D	D	B	D
	11	3	38	A	A	D	D	C	D
	8	4	37	A	A	D	D	C	D
	5	5	36	A	A	D	B	C	
	2	6	35	A	A	D	D	B	D
	7	7	34	A	D	D	B	B	D
	6	8	33	A	C	A	B	A	D
	9	10	31	A	C	D	B	C	A
	4	12	29	A	A	D	D	C	D
	12	12	29	A	C	D	D	A	A
	3	16	25	A	D	D	D	A	D

Figure 6.2: Image of scores and responses of 6 items from 12 students.

demonstrate how to evaluate my exam items with these indices of difficulty and discrimination. I have intentionally selected items with a variety of desirable (and undesirable) characteristics.

**Item 5** A grouping variable such as men or women that uses dummy coding of 1 and 0 to categorize the groups is an example of \_\_\_\_\_ scaling.

- a) Nominal
- b) Ordinal
- c) Interval
- d) Ratio

If we wanted to hand-calculate the index of discrimination for Item #5, we find that 3 people (100%) in the upper group selected the correct answer and 3 people (100%) in the lower group selected the correct answer:  $3 - 3 = 0$ . If you prefer percentages:  $100\% - 100\% = 0\%$ . This means there is no discrimination in performance of the upper and lower performing groupings.

Older scoring systems (e.g., Scantron) used to provide this information.

Considering what we have learned already, Item #5 is:

- too easy
- does not discriminate between upper and lower performance
- *Yes, there is more data on here, but we will save it for the next level of review...just a few moments.*

**Item 11** The term “grade inflation” has frequently been applied to describe the distribution of grades in graduate school. Which of the following best describes this distribution.

- a) negatively skewed
- b) uniform/rectangular

<b>Item 5</b>	Answer: A Point Biserial: 0.00	Correct Responses: Total Group: Upper 27% of Group: Lower 27% of Group:
Distractor Analysis - * indicates correct answer:		
	*A      B      C      D      E	
Pt-Biser.:	0.00      0.00      0.00      0.00      0.00	
Frequency%:	100.00      0.00      0.00      0.00      0.00	

Figure 6.3: Scantron image of item analysis for exam item #5

- c) positively skewed and leptokurtic
- d) uniform and platykurtic

For Item #11, 2 people (~66%) from the upper group selected the correct answer, 1 person (~33%) from the lower group selected the correct answer. Thus, the U-L was +1 (+33%) and the item is working in the proper direction.

<b>Item 11</b>	Answer: A Point Biserial: 0.44	Correct Responses: Total Group: 50.00% Upper 27% of Group: 66.67% Lower 27% of Group: 33.33%
Distractor Analysis - * indicates correct answer:		
	*A      B      C      D      E	
Pt-Biser.:	0.44      0.00      -0.15      -0.39      0.00	
Frequency%:	50.00      0.00      33.33      16.67      0.00	

Figure 6.4: Scantron image of item analysis for exam item #11

Considering what we have learned already, Item #11 is:

- difficult (50% overall selected the correct item)
- does discriminate between upper and lower performance, with more individuals in the upper groups selecting the correct answer than in the lower group

**Item 19** All distributions of Z-scores will have the identical

- a) Mean
- b) Variance
- c) Standard deviation
- d) All of the above

<b>Item 19</b>	Answer: D Point Biserial: -0.06	Correct Responses: Total Group: 91.87% Upper 27% of Group: 100.00% Lower 27% of Group: 100.00%
Distractor Analysis - * indicates correct answer:		
A	B	C
Pt-Biser.: 0.06	0.00	0.00
Frequency%: 8.33	0.00	91.67
D		
E		
Pt-Biser.: 0.06 0.00 -0.06 0.00		
Frequency%: 8.33 0.00 91.67 0.00		
** Review item(s): A		

Figure 6.5: Scantron image of item analysis for exam item #19

Hand calculation: Upper = 3 (100%), Lower = 3 (100%). Difference = 0.

Considering what we have learned already, Item #19 is:

- somewhat easy (92% overall selected the correct item)
- using the U - L discrimination index, it does not discriminate between upper and lower performance

**Item 21** The most appropriate score for comparing scores across two or more distributions (e.g., exam scores in math and art classes) is the:

- a) mean
- b) percentile rank
- c) raw score
- d) z-score

Hand calculation: Upper = 2 (66%), Lower = 3 (100%). Difference = -33%. This item is upside down. This is different than the Scantron snip below because uppers and lowers were likely calculated on exam total that included subjectively scored items (essays; and I no longer have that data).

<b>Item 21</b>	Answer: D Point Biserial: -0.13	Correct Responses: Total Group: 58.33% Upper 27% of Group: 66.67% Lower 27% of Group: 66.67%
Distractor Analysis - * indicates correct answer:		
A	B	C
Pt-Biser.: 0.00	0.13	0.00
Frequency%: 0.00	41.67	0.00
D	E	
Pt-Biser.: 0.00 -0.13 0.00		
Frequency%: 0.00 41.67 0.00		
** Review item(s): B		

Figure 6.6: Scantron image of item analysis for exam item #21

Considering what we have learned already, Item #21 is:

- somewhat difficult (58% overall selected the correct item)
- on the basis of the hand-calculations it does not discriminate between uppers and lowers

**Item 37** Of the following, what statement best describes  $r^2 = .49$

- a) strong positive correlation

- b) strong positive or negative correlation
- c) weak positive or negative correlation
- d) weak negative correlation

Hand calculation: Upper = 2 (66%), Lower = 0 (0%). Difference = 66%.

<b>Item 37</b>	Answer: B Point Biserial: 0.48	Correct Responses: Total Group: 33.33% Upper 27% of Group: 66.67% Lower 27% of Group: 0.00%		
Distractor Analysis - * indicates correct answer:				
A	*B	C	D	E
Pt-Biser.: -0.52	0.48	0.00	0.00	0.00

Figure 6.7: Scantron image of item analysis for exam item #37

Considering what we have learned already, Item #37 is:

- very difficult (33% overall selected the correct item)
- on the basis of the hand-calculations, this completely discriminates the uppers from the lowers)

**Item 38** When there are no ties among ranks, what is the relationship between the Spearman rho ( $\rho$ ) and the Pearson r ( $r$ )?

- a)  $\rho = r$
- ii)  $\rho > r$
- a)  $\rho < r$
- b) no relationship

Hand calculation: Upper = 1 (33%), Lower = 1 (33%). Difference = 0%.

<b>Item 38</b>	Answer: A Point Biserial: -0.21	Correct Responses: Total Group: 25.00% Upper 27% of Group: 33.33% Lower 27% of Group: 33.33%		
Distractor Analysis - * indicates correct answer:				
'A	B	C	D	E
Pt-Biser.: -0.19	0.00	0.00	0.19	0.00

Figure 6.8: Scantron image of item analysis for exam item #38

Considering what we have learned already, Item #21 is:

- very difficult (25% overall selected the correct item)
- on the basis of the hand-calculations, this does not discriminate the uppers from the lowers

## 6.6 In the *psych* Package

Using the *score.multiple.choice()* function in the *psych* package. Documentation is pp. 85-86 in <http://personality-project.org/r/overview.pdf>

A multiple choice exam presumes that there is one correct response. We start with a dataset that records the students' responses. It *appears* that the psych package requires these responses to be numerical (rather than A, B, C, D).

```
# For portability of the lesson, I hand-entered the exam score data.
# Variables are items (not students), so the entry is the 41 items
# for the 12 students
Item1 <- c(1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1)
Item2 <- c(4, 4, 1, 4, 4, 4, 4, 4, 4, 4, 4)
Item3 <- c(1, 1, 4, 1, 1, 1, 1, 3, 1, 1, 1)
Item4 <- c(2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2)
Item5 <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item6 <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item7 <- c(3, 3, 4, 4, 3, 3, 3, 3, 3, 3, 2)
Item8 <- c(1, 2, 2, 4, 2, 2, 2, 1, 4, 2, 2)
Item9 <- c(1, 1, 4, 4, 1, 4, 1, 1, 1, 1, 4)
Item10 <- c(3, 3, 3, 2, 3, 2, 3, 2, 2, 3, 3)
Item11 <- c(1, 1, 4, 1, 1, 3, 4, 1, 3, 3, 1, 3)
Item12 <- c(2, 1, 2, 4, 2, 2, 2, 2, 2, 2, 2)
Item13 <- c(2, 2, 3, 3, 2, 2, 2, 2, 2, 2, 1)
Item14 <- c(2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2)
Item15 <- c(2, 1, 1, 3, 2, 4, 2, 2, 2, 2, 4, 2)
Item16 <- c(2, 2, 2, 4, 4, 2, 2, 2, 4, 2, 2, 1)
Item17 <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item18 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item19 <- c(4, 4, 4, 4, 1, 4, 4, 4, 4, 4, 4)
Item20 <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item21 <- c(2, 4, 4, 4, 2, 2, 2, 4, 2, 4, 4, 4)
Item22 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 1)
Item23 <- c(3, 3, 2, 3, 3, 3, 3, 2, 3, 3, 2)
Item24 <- c(3, 3, 1, 3, 3, 2, 3, 3, 3, 3, 1)
Item25 <- c(2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2)
Item26 <- c(4, 4, 4, 4, 4, 4, 4, 1, 4, 4, 1)
Item27 <- c(4, 4, 1, 4, 4, 4, 4, 4, 4, 4, 4)
Item28 <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item29 <- c(1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Item30 <- c(2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)
Item31 <- c(1, 1, 1, 2, 1, 1, 3, 1, 1, 1, 1, 2)
Item32 <- c(1, 1, 3, 1, 1, 1, 3, 1, 1, 1, 1, 1)
Item33 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item34 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item35 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item36 <- c(2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)
```

```

Item37 <- c(2, 2, 1, 3, 3, 1, 2, 3, 3, 2, 3, 1)
Item38 <- c(1, 4, 4, 4, NA, 4, 4, 4, 1, 4, 4, 1)
Item39 <- c(3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item40 <- c(3, 3, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3)
Item41 <- c(2, 1, 2, 2, 4, 4, 2, 2, 4, 4, 2)

exam <- data.frame(Item1, Item2, Item3, Item4, Item5, Item6, Item7, Item8,
  Item9, Item10, Item11, Item12, Item13, Item14, Item15, Item16, Item17,
  Item18, Item19, Item20, Item21, Item22, Item23, Item24, Item25, Item26,
  Item27, Item28, Item29, Item30, Item31, Item32, Item33, Item34, Item35,
  Item36, Item37, Item38, Item39, Item40, Item41)

```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do).

```

# write the simulated data as a .csv write.table(exam,
# file='exam.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file exam <- read.csv
# ('exam.csv', header = TRUE)

```

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saverDS(exam, 'exam.rds') bring back the simulated dat from an
# .rds file exam <- readRDS('exam.rds')

```

We create a key of the correct answers.

```
exam.keys <- c(1,4,1,2,1,1,3,2,1,3,1,2,2,2,2,2,1,3,4,1,4,3,3,3,1,4,4,1,1,2,1,1,3,3,3,2,2,1,3,
```

We then insert that key into the *psych* package’s *score.multiple.choice()* function.

```
results <- psych::score.multiple.choice(exam.keys, exam, score = TRUE,
  short = FALSE, skew = TRUE)
```

```
Warning in cor(items, scores, use = "pairwise"): the standard deviation is zero
```

```
results
```

Call: NULL

```
(Unstandardized) Alpha:
[1] 0.73
```

```
Average item correlation:
[1] 0.06
```

## item statistics

	key	1	2	3	4	miss	r	n	mean	sd	skew	kurtosis	se
Item1	1	0.92	0.00	0.00	0.08	0.00	0.65	12	0.92	0.29	-2.65	5.48	0.08
Item2	4	0.08	0.00	0.00	0.92	0.00	0.65	12	0.92	0.29	-2.65	5.48	0.08
Item3	1	0.83	0.00	0.08	0.08	0.00	0.34	12	0.83	0.39	-1.57	0.53	0.11
Item4	2	0.00	0.92	0.08	0.00	0.00	-0.11	12	0.92	0.29	-2.65	5.48	0.08
Item5	1	1.00	0.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item6	1	1.00	0.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item7	3	0.00	0.08	0.67	0.25	0.00	0.72	12	0.67	0.49	-0.62	-1.74	0.14
Item8	2	0.17	0.67	0.00	0.17	0.00	-0.13	12	0.67	0.49	-0.62	-1.74	0.14
Item9	1	0.67	0.00	0.00	0.33	0.00	0.81	12	0.67	0.49	-0.62	-1.74	0.14
Item10	3	0.00	0.33	0.67	0.00	0.00	0.18	12	0.67	0.49	-0.62	-1.74	0.14
Item11	1	0.50	0.00	0.33	0.17	0.00	0.42	12	0.50	0.52	0.00	-2.16	0.15
Item12	2	0.08	0.83	0.00	0.08	0.00	0.17	12	0.83	0.39	-1.57	0.53	0.11
Item13	2	0.08	0.75	0.17	0.00	0.00	0.85	12	0.75	0.45	-1.01	-1.04	0.13
Item14	2	0.00	0.92	0.08	0.00	0.00	-0.04	12	0.92	0.29	-2.65	5.48	0.08
Item15	2	0.17	0.58	0.08	0.17	0.00	0.32	12	0.58	0.51	-0.30	-2.06	0.15
Item16	2	0.08	0.67	0.00	0.25	0.00	0.40	12	0.67	0.49	-0.62	-1.74	0.14
Item17	1	1.00	0.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item18	3	0.00	0.00	1.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item19	4	0.08	0.00	0.00	0.92	0.00	0.04	12	0.92	0.29	-2.65	5.48	0.08
Item20	1	1.00	0.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item21	4	0.00	0.42	0.00	0.58	0.00	-0.19	12	0.58	0.51	-0.30	-2.06	0.15
Item22	3	0.08	0.00	0.92	0.00	0.00	0.34	12	0.92	0.29	-2.65	5.48	0.08
Item23	3	0.00	0.25	0.75	0.00	0.00	0.71	12	0.75	0.45	-1.01	-1.04	0.13
Item24	3	0.17	0.08	0.75	0.00	0.00	0.61	12	0.75	0.45	-1.01	-1.04	0.13
Item25	1	0.00	0.92	0.08	0.00	0.00	NA	12	0.00	0.00	NaN	NaN	0.00
Item26	4	0.17	0.00	0.00	0.83	0.00	0.34	12	0.83	0.39	-1.57	0.53	0.11
Item27	4	0.08	0.00	0.00	0.92	0.00	0.65	12	0.92	0.29	-2.65	5.48	0.08
Item28	1	1.00	0.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item29	1	0.92	0.00	0.08	0.00	0.00	-0.11	12	0.92	0.29	-2.65	5.48	0.08
Item30	2	0.00	1.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item31	1	0.75	0.17	0.08	0.00	0.00	0.41	12	0.75	0.45	-1.01	-1.04	0.13
Item32	1	0.83	0.00	0.17	0.00	0.00	0.45	12	0.83	0.39	-1.57	0.53	0.11
Item33	3	0.00	0.00	1.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item34	3	0.00	0.00	1.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item35	3	0.00	0.00	1.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item36	2	0.00	1.00	0.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item37	2	0.25	0.33	0.42	0.00	0.00	0.49	12	0.33	0.49	0.62	-1.74	0.14
Item38	1	0.27	0.00	0.00	0.73	0.08	-0.07	11	0.27	0.47	0.88	-1.31	0.14
Item39	3	0.00	0.00	1.00	0.00	0.00	NA	12	1.00	0.00	NaN	NaN	0.00
Item40	3	0.00	0.00	0.92	0.08	0.00	0.65	12	0.92	0.29	-2.65	5.48	0.08
Item41	4	0.08	0.58	0.00	0.33	0.00	0.40	12	0.33	0.49	0.62	-1.74	0.14

```
# short=FALSE allows us to produce scores; we will use these later in
# some IRT analyses names(results)
```

The first screen of output provides an alpha. In this context, *alpha* should tell us the consistency of getting answers right or wrong. Technically, the alpha is reduced to a KR-20 (Kuder Richardson 20). We interpret it the same. Alpha is directly effected by:

- *interitem correlations* among the items – a large number of positive correlations between items increases alpha
- test length – more items produce higher reliability (all things else equal)
- test content – the more diverse/broad, the lower the reliability coefficient

In the context of the classroom, reliabilities above .70 are probably adequate and above .80 are good. Reliabilities below .60 suggest that items should be investigated and additional measures (tests, homework assignments) should be included in assigning grades.

Focus instead on the second screen of output.

**key** indicates which answer was correct.

**1, 2, 3, 4** (there would be as many as there are options in the multiple choice exam) provide a *distractor analysis* by indicating the percentage of time that answer was chosen. For item 1, option 1 was correct, and it was chosen 92% of the time. No individuals chose options 2 or 3. Option 4 was chosen 8% of the time.

**miss** indicates how many times the item was skipped.

*r* is a point-biserial correlation with a dichotomous correct/incorrect correlated with the continuously scaled total scale score. Positively scored items let us know that the item is working in the proper direction; the students who got the item correct, did better on the overall total score and vice versa.

- one of the best indicators of an items ability to *discriminate* (hence, **item discrimination**) among the criterion assessed on the test
- it is important to investigate those with values close to zero (no relation between item performance with overall test performance) and those with negative values (meaning that those who had the correct answer on the item were those who scored lower on the exam).

*n* tells us how many participants completed the item (this would necessarily be the inverse of “miss”).

*mean* repeats the proportion of individuals who scored correctly; it would be the same as the percentage in the item keyed as the correct one. This is an indication of **item difficulty**.

*sd* gives an indication of the variability around that mean

It is important to look at the *r* and *mean* columns, together to understand the degree of difficulty and how well each item is discriminating between performance levels.

*skew* can provide an indication of ceiling and floor effects.

If a score has a significant negative skew (long tail to the left), then there may be a piling up of items at the upper end of the scale. This would indicate an *insufficient ceiling* and make it more difficult to discriminate among differences among the higher performers.

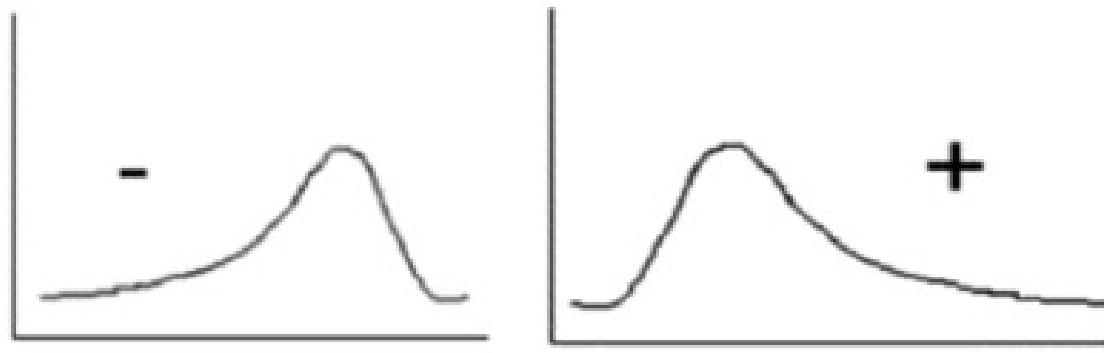


Figure 6.9: Image of two graphs illustrating positive and negative skew

If a score has a significant positive skew (long tail to the right), then there may be a piling up of items at the low end, indicating an *insufficient floor*. That is, it lacks the ability to discriminate between poorer performers.

How do you tell what is significant?

A general rule of thumb says that anything greater or less than the absolute value of 1.0 is significantly skewed. A formal z-test can be conducted this way:  $z_{skewness} = \frac{S-0}{SE_{skewness}}$

In our exam dataset, -2.65 is the most extremely negatively skewed item and its  $se = 0.08$ .

```
-2.65/0.08
```

```
[1] -33.125
```

Considering that anything greater than  $\pm 1.96$  is statistically significant, it is safe to say that this item has an insufficient ceiling.

What about the items with -0.30 ( $se = 0.15$ )?

```
-.30/.15
```

```
[1] -2
```

This is not as extreme (and recall my  $N = 12$ , so I should probably look up a critical  $t$  value), but there is still some question about whether my exam items can discriminate among high performers.

Please note, because these distributions are *dichotomous* (correct/incorrect) they will never be normally distributed, but, like the difficulty index, they give another glimpse of the ability to discriminate.

Before we look at the specific exam items and their output from the scoring function, let me introduce you to the features of the psych package that draw from *item response theory* (IRT).

### 6.6.1 A Mini-Introduction to IRT

To recap – at the instructional level, the combination of percent passing (mean) and point-biserial correlation (discrimination index) is status quo for evaluating/improving the items.

The *psych* package draws from its IRT capacity to conduct distractor analysis. IRT models individual responses to items by estimating individual ability (*theta*) and item difficulty (*diff*) parameters.

In these graphs, theta is on the X axis. Theta is the standard unit of the IRT model that represents the level of the domain being measured. Like a z-score, a theta unit of “1” is the SD of the calibrated sample.

The pattern of responses to multiple choice ability items can show that some items have poor distractors. This may be done by using the *irt.responses* function. A good distractor is one that is negatively related to ability.

As we look at each of the exam items, we will look at the *psych* input from the *scoring* function as well as use the *results* objects to create the IRT graphs.

**Item 5** A grouping variable such as men or women that uses dummy coding of 1 and 0 to categorize the groups is an example of \_\_\_\_\_ scaling.

- a) Nominal
- b) Ordinal
- c) Interval
- d) Ratio

Mean = 1.0 (much too easy),  $r = \text{NA}$ , Distractors: 1.00 0.00 0.00 0.00, skew = -2.65

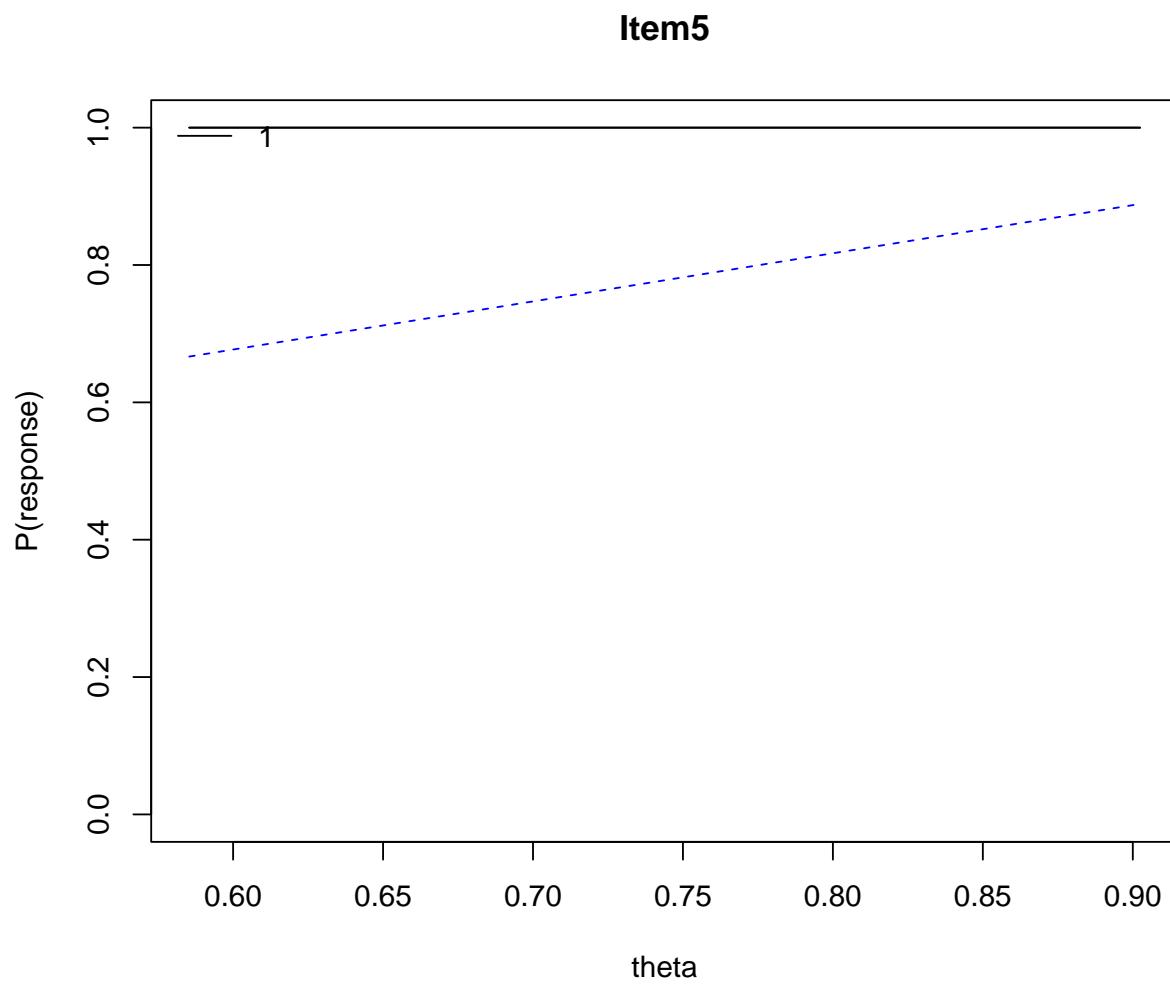
```
# irt.responses(scores$scores, exam[5], breaks = 2)
psych::irt.responses(results$scores, exam[5], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



With Item #5, 100% responded correctly (the flat, solid line at the top); there is not much to see.

**Item 11** The term “grade inflation” has frequently been applied to describe the distribution of grades in graduate school. Which of the following best describes this distribution.

- a) negatively skewed
- b) uniform/rectangular
- c) positively skewed and leptokurtic
- d) uniform and platykurtic

Mean = .50,  $r = .42$ , Distractors: 0.50 0.00 0.33 0.17, skew = 0.00

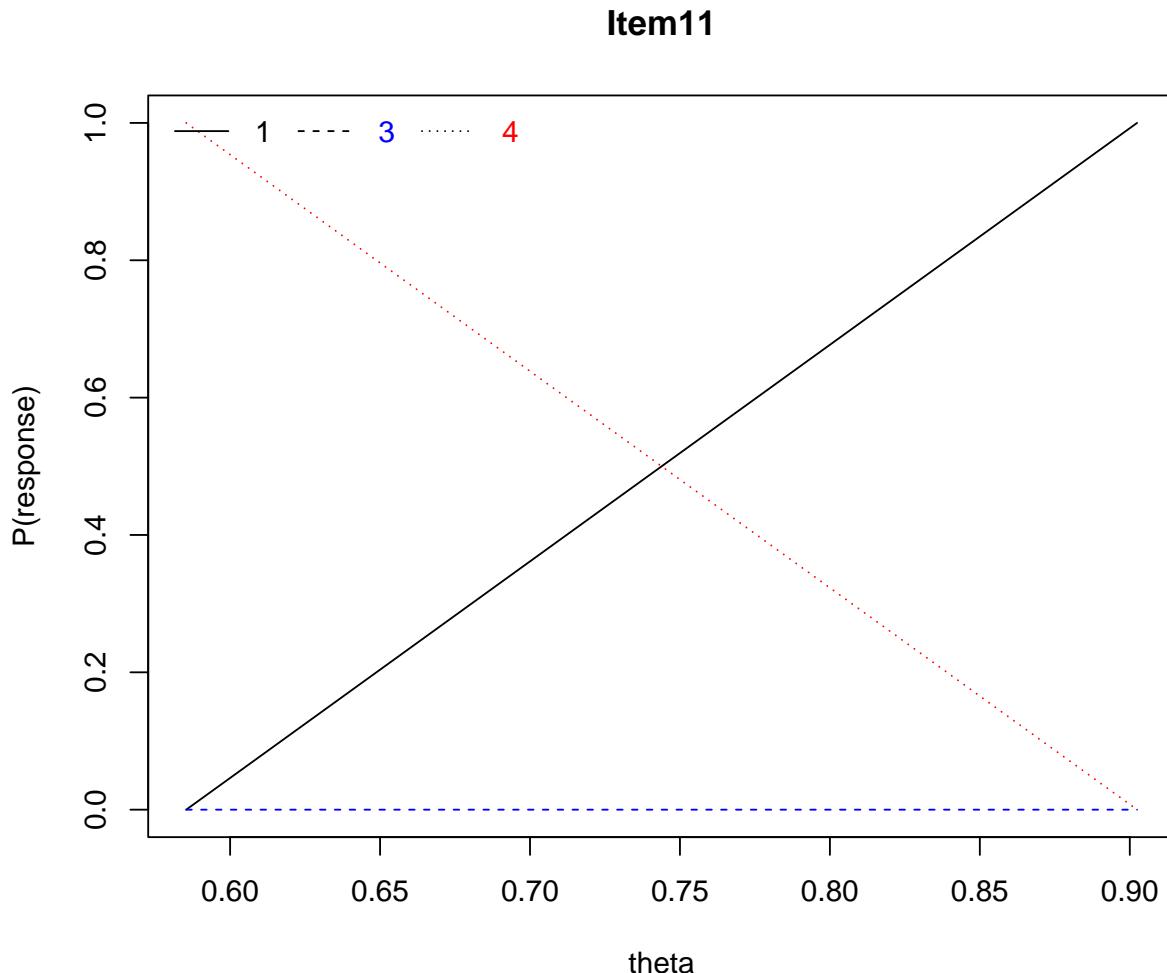
```
psych::irt.responses(results$scores, exam[11], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



With Item #11, there is a positive relationship between 1/A (correct answer) and ability (theta), no relationship between 3/C and ability, and a negative relationship between 4/D and ability (indicating that 4/D is a good distractor). These map onto each of the point-biserial correlations associated with the distractors in the Scantron output.

**Item 19** All distributions of Z-scores will have the identical

- a) Mean
- b) Variance

- c) Standard deviation
- d) All of the above

Mean = .92,  $r = .04$ , Distractors: 0.08 0.00 0.00 0.92 , skew = -2.65

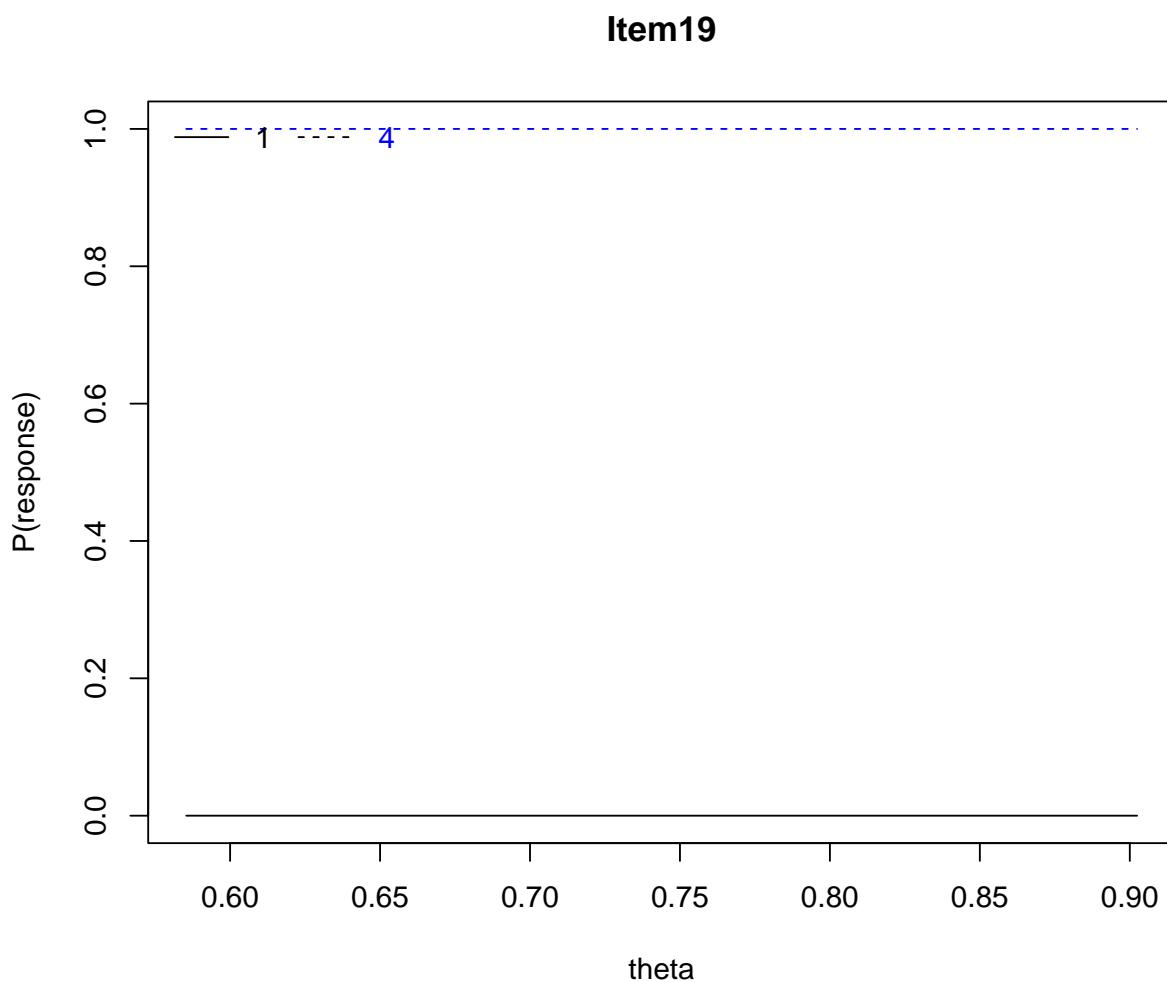
```
psych::irt.responses(results$scores, exam[19], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



Item #19 shows rather flat (no relationship) relations with ability for the correct item and the lone distractor.

**Item 21** The most appropriate score for comparing scores across two or more distributions (e.g., exam scores in math and art classes) is the:

- a) mean
- b) percentile rank
- c) raw score
- d) z-score

Mean = .58,  $r = -.19$ , Distractors: 0.00 0.42 0.00 0.58, skew = -0.30

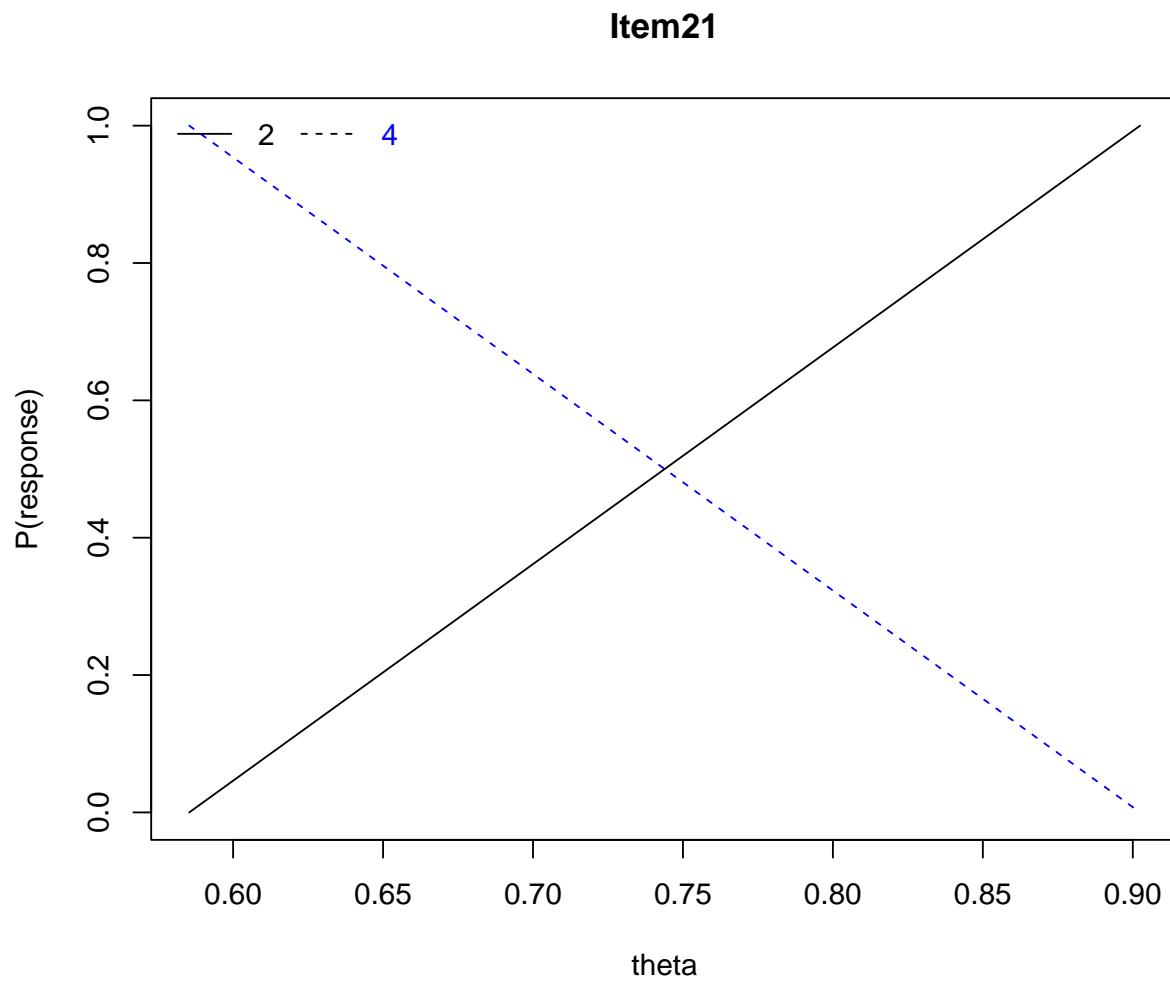
```
psych::irt.responses(results$scores, exam[21], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



For Item #21, a positive relationship between the WRONG answer (2/B) and ability (theta) and a negative relationship between 4/D (incorrect answer) and ability. This makes sense as the point biserial for the overall item was 0-.13.

**Item 37** Of the following, what statement best describes  $r^2 = .49$

- a) strong positive correlation
- b) strong positive or negative correlation
- c) weak positive or negative correlation
- d) weak negative correlation

Mean = .33,  $r = .49$ , Distractors: 0.25 0.33 0.42 0.00, skew = .62

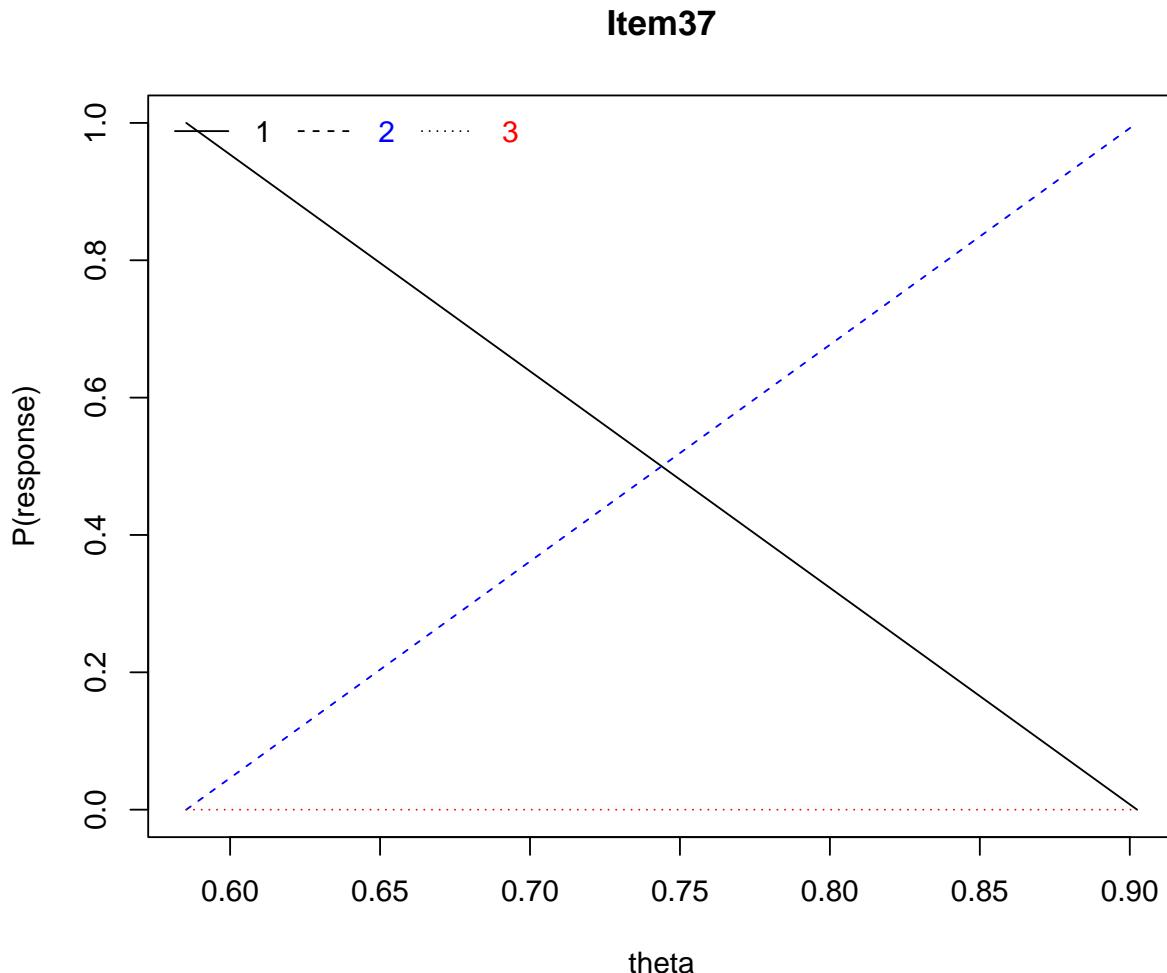
```
psych::irt.responses(results$scores, exam[37], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



For Item #37, a negative relation between endorsing 1/A and ability (a good distractor). No relationship with ability for endorsing 3/C. A positive relation with ability for those endorsing 2/B (correct answer).

**Item 38** When there are no ties among ranks, what is the relationship between the Spearman rho ( $\rho$ ) and the Pearson r ( $r$ )?

- a)  $\rho = r$
- ii)  $\rho > r$

- a)  $\rho < r$
- b) no relationship

Mean = .27,  $r = -.07$ , Distractors: 0.27 0.00 0.00 0.73, skew = .68

Notice anything else that's funky about Item #38?

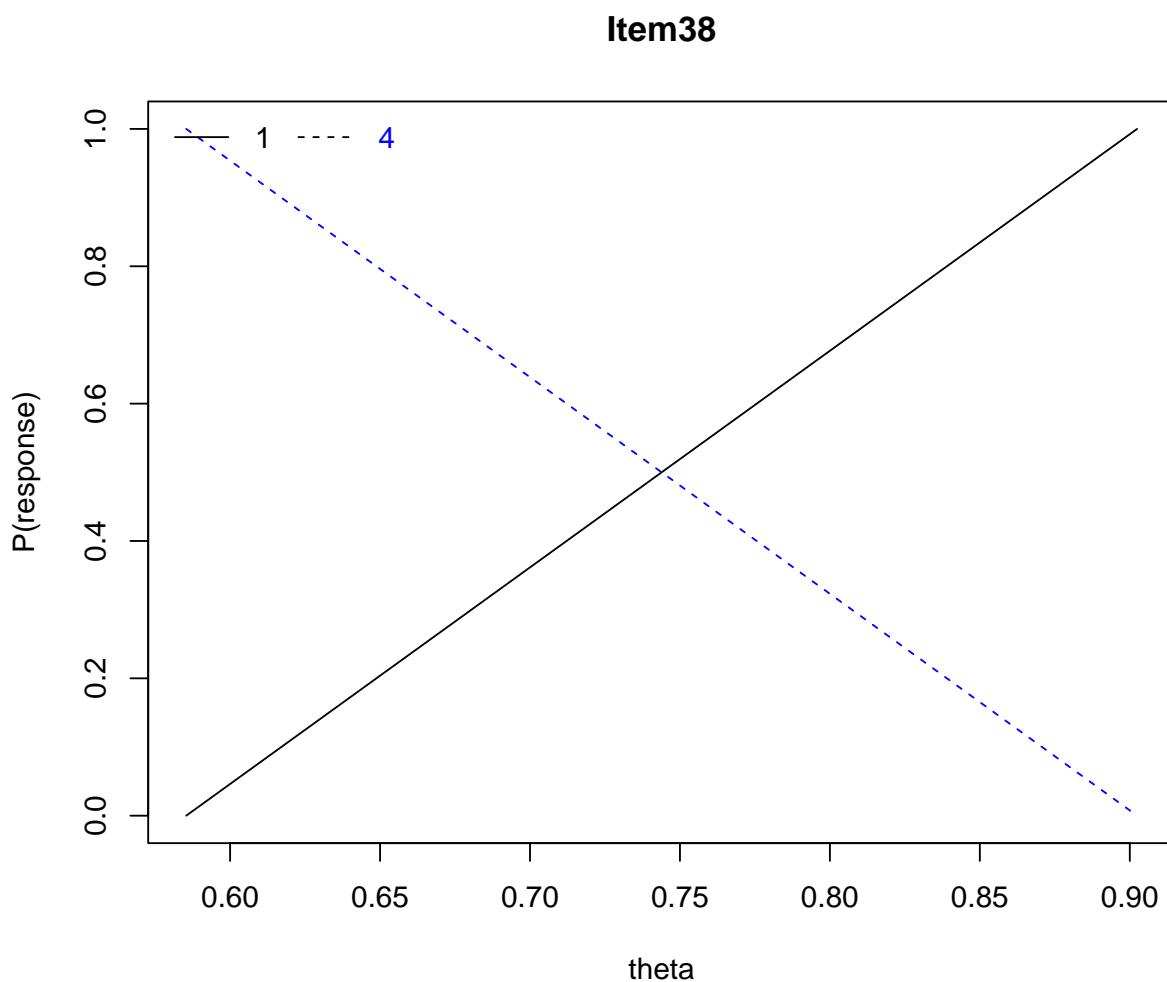
```
psych::irt.responses(results$scores, exam[38], breaks = 2)
```

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)

Number of categories should be increased in order to count frequencies.

Warning in rbind(items, dummy): number of columns of result is not a multiple of vector length (arg 1)



## 6.7. CLOSING THOUGHTS ON DEVELOPING MEASURES IN THE EDUCATION/ACHIEVEMENT CONTEXT

For Item #38, there is a positive relationship with ability for endorsing 1/A (correct answer) and a negative relationship with ability for 4/D (incorrect answer).

Regarding overall test characteristics

3340 Final Exam Detailed Item Analysis Report On Final Exam Version A Version A				
Course #:		Instructor:	Bikos	
Course Title:		Description:		
Day/Time:		Term/Year:		
Total Possible Points:	56.00	Median Score:	49.00	Mean Score: 46.83
Student in this group:	12	Mode Score:	53.00	Variance: 37.14
Standard Deviation:	6.09	Highest Score:	53.00	Lowest Score: 35.00
Measure of Skewness:	-0.64	Degree of Kurtosis:	-1.07	Reliability Coefficient (KR20): 0.77

## 6.7 Closing Thoughts on Developing Measures in the Education/Achievement Context

Item analysis tends to be an assessment of *reliability*. However, in the context of educational assessment and achievement exams, there are also *validity* issues.

**Content validity** is concerned with whether or not the scale adequately represents the entirety of the *domain* to be assessed.

In educational and achievement contexts, this is often accomplished with a *table of specifications*. I introduced this in the [Validity lesson](#). As a refresher, I will include another example – imagining that I am going to write a quiz or short exam based on the learning objectives of this, single, lesson. There are a number of different ways to organize the types of knowledge that is being assessed. Since the American Psychological Association (and others) work in “KSAs” (knowledge, skills, attitudes) in their accreditation standards, I will use those.

In creating a table of specifications, we start with the learning objectives. Then we decide what type of items to write and what type of performance level they satisfy. This helps us ensure that all learning objectives are proportionately covered, using a variety of assessment approaches. Otherwise, we might be tempted to include the items that come easily to us or that are from our favorite topics. Personally, I find that when I work on the exam, and am informed by the learning objectives and table of specifications, I find myself tinkering with all three. I am inclined to believe that this results in an ever-increasingly-improved pedagogy.

### Table of Specifications

Learning Objectives	Knowledge	Skills	Attitudes	% of test
Provide a rationale for why having a <i>test bank</i> might be a good idea.			1 item	30%
Describe the effects of skewness on the interpretation of exam results.	2 items			10%
Evaluate the quality of a multiple choice item on the basis of item difficulty, correlation, and discrimination.		5 items		25%

Learning Objectives	Knowledge	Skills	Attitudes	% of test
Discuss the challenges of identifying an <i>ideal</i> difficulty level for test items. Further elaborate how guessing, speeded tests, interitem correlations, and the purposes of the test influence the <i>ideal difficulty</i> .	2 items		1 item	35%
TOTALS	4 items	5 items	2 items	100%

There are a variety of free resources that help with this process. Below are some that I find helpful:

- [Bloom's Taxonomy Verbs](#), freely available from Fractus Learning.
- [The Bloom's Taxonomy Verbs Poster for Teachers](#)
- If you have “writer’s block” for writing objectives, here is a [learning outcome generator](#) that may help get you started.
- From APA’s Education Directorate, [Guidance for Writing Behavioral Learning Objectives](#). The APA Guidance really emphasizes key components of well-written behavioral leaning objectives. These include:
  - **observable and measurable**, using action verbs that describe measureable behaviors. The APA CE office disallows the use of “understand” as an action verb,
  - statements that clearly describe what the learner will know or be able to do **as a result** of having participated,
  - focused on the learner and learning (as opposed to what the trainer is doing or leading),
  - appropriate in breadth (not too few or too many)

**Takeaway message:** Together, mapping out exam coverage in a table of specifications PLUS item analysis (difficulty/discrimination) can be powerful tools in educational assessment.

## 6.8 Practice Problems

For this particular lesson, I think some of the most meaningful practice comes from multiple choice and true/false exams that occur in your life. If you are in a class, see if your instructor is willing to share item analysis information that they have received. Learning management systems like Canvas, automatically calculate these.

If you are an instructor, calculate and review item analysis data on your own items. Think about how you might improve items between exams and cconsider how the dificulty and discrimination capacity of the item changes.

# Chapter 7

## Item Analysis for Likert Type Scale Construction

### [Screencasted Lecture Link](#)

The focus of this lecture is on item analysis for surveys. We use information about alpha coefficients and item-total correlations (within and across subscales) to help assess what we might consider to be *within-scale convergent and discriminant validity* (although we tend to think of it as an assessment of reliability).

### 7.1 Navigating this Lesson

There is about 45 minutes of lecture. If you work through the materials with me it would be plan for an additional hour.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

#### 7.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Define the corrected item-total correlation and compare it to an item-total correlation.
- List the preliminary steps essential for scale construction, beginning with item development.
- Name the type(s; e.g., reliability, validity) of psychometric evaluation that item analytic procedures assess.
- Identify threats to the interpretation of item-total correlations and alpha coefficients.
- Make decisions about item retention, deletion, and revision that balances statistical output with construct definitions.

### 7.1.2 Planning for Practice

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. For this lesson, please locate item-level data for a scale that has the potential for at least two subscales and a total-scale score. Ideally, the data you utilized in one or more of the prior lessons (e.g., changing the random seed in the lesson data, downloading the data from the *ReCentering Psych Stats* survey, or data you found elsewhere) will allow you to continue with these analyses. Then, please examine the following:

- Produce alpha coefficients, average inter-item correlations, and corrected item-total correlations for the total and subscales, separately.
- Produce correlations between the individual items of one subscale and the subscale scores of all other scales.
- Draft an APA style results section with an accompanying table.

In my example there were only two subscales. If you have more, you will need to compare each subscale with all the others. For example, if you had three subscales: A, B, C, you would need to compare A/B, B/C, and A/C.

### 7.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

- Green & Salkind (2017). Lesson 38: Item analysis using the reliability Procedure. In S.B. Green and N.J. Salkind's, "Using SPSS for Windows and Macintosh: Analyzing and understanding data (8th ed). New York: Pearson.
  - Even though the operation of the chapter uses SPSS, the narration of the "what" and "why" of item analysis is clear and concise. Further, I have not found another chapter (not even in psychometrics texts) that addresses this as completely.
- Szymanski, D. M., & Bissonette, D. (2020). Perceptions of the LGBTQ College Campus Climate Scale: Development and psychometric evaluation. *Journal of Homosexuality*, 67(10), 1412–1428. <https://doi.org/10.1080/00918369.2019.1591788>
  - The research vignette for this lesson.

### 7.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(tidyverse)){install.packages('tidyverse')}
# if(!require(MASS)){install.packages('MASS')}
```

```
# if(!require(psych)){install.packages('psych')}
# if(!require(apaTables)){install.packages('apaTables')}
# if(!require(sjstats)){install.packages('sjstats')}
# if(!require(qualtrics)){install.packages('qualtrics')}
```

## 7.2 Introducing Item Analysis for Survey Development

Item analysis can be used to help determine which items to include and exclude from a scale or subscale. The goal is to select a set of items that yields a summary score (total or mean) that is strongly related to the construct identified and defined in the scale.

- Item analysis is somewhat limiting because we usually cannot relate our items to a direct (external) measure of a construct to select our items.
- Instead, we *trust* (term used lightly) that the items we have chosen, together, represent the construct and we make decisions about the relative strength of each item's correlation to the total score.
- This makes it imperative that we look to both statistics and our construct definition (e.g., how well does each item map onto the construct definition)

If this is initial scale development, the researchers are wise to write more items than needed so that there is flexibility in selecting items with optimal functioning. Szymanski and Bissonette [2020] do this. Their article narrates how they began with 36 items, narrowed it to 24, and – on the basis of subject matter expertise and peer review – further narrowed it to 10. The reduction of additional items happened on the basis of exploratory factor analysis.

### 7.2.1 Workflow for Item Analysis

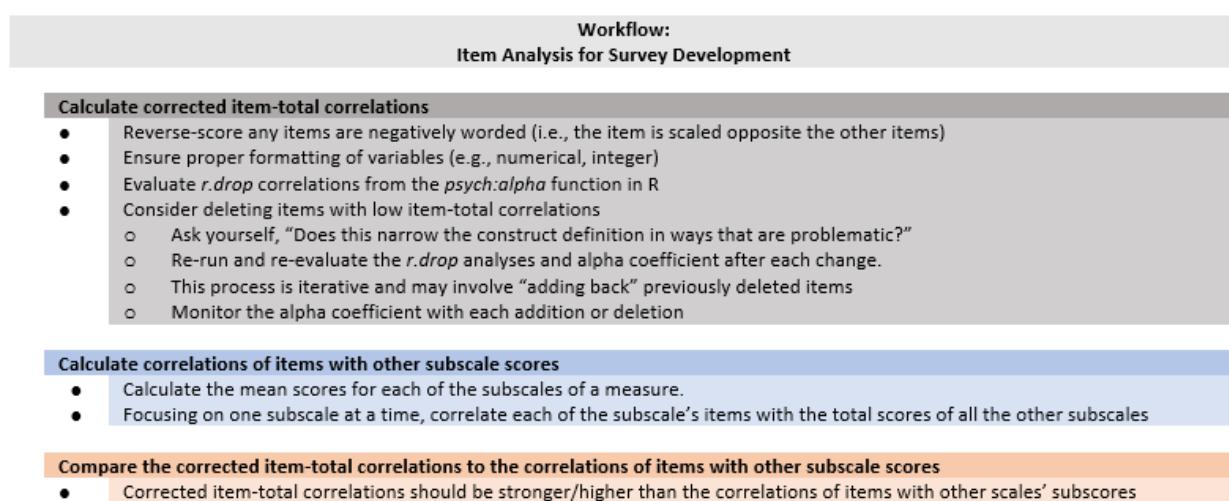


Figure 7.1: Image of workflow for item analysis for survey development.

Step I: Calculate corrected item-total correlations. This involves:

- Reverse-scoring items that are negatively worded.
- Ensuring proper formatting of variables (i.e., numerical and integer formats).
- Evaluating the corrected item-total correlations (“*r.drop*” in the *psych::alpha* function)
- Consider deleting items with low item-total correlations.
  - Consider the how deleting items might create too narrow of a construct definition. If so, hesitate before deleting.
  - Re-run and re-evaluate the *r.drop* values and alpha coefficients after each change.
  - This is an interative process and may involved “adding back” previously deleted items.
- Calculate correlations of items with other subscale scores.
  - Calculate the mean scores for each of the subscales of a measure.
  - Focusing on one subscale at a time, correlate each of the subscale’s items with the total score of all the other subscales.
- Compare the corrected item-total correlations to the correlations of items with other subscale scores.
  - The correted item-total correlations should be stronger/higher than the correlations of items with other scales’ subscores.

### 7.3 Research Vignette

The research vignette for this lesson is the development and psychometric evaluation of the Perceptions of the LGBTQ College Campus Climate Scale [Szymanski and Bissonette, 2020]. The scale is six items with responses rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores indicate more negative perceptions of the LGBTQ campus climate. Szymanski and Bissonette [2020] have suggested that the psychometric evaluation supports using the scale in its entirety or as subscales. Each item is listed below with its variable name in parentheses:

- College response to LGBTQ students:
  - My university/college is cold and uncaring toward LGBTQ students. (cold)
  - My university/college is unresponsive to the needs of LGBTQ students. (unresponsive)
  - My university/college provides a supportive environment for LGBTQ students. (unsupportive)
    - \* this item must be reverse-scored
- LGBTQ Stigma:
  - Negative attitudes toward LGBTQ persons are openly expressed on my university/college campus. (negative)
  - Heterosexism, homophobia, biphobia, transphobia, and cissexism are visible on my university/college campus. (heterosexism)
  - LGBTQ students are harassed on my university/college campus. (harassed)

A preprint of the article is available at ResearchGate. Below is the script for simulating item-level data from the factor loadings, means, and sample size presented in the published article.

Because data is collected at the item level (and I want this resource to be as practical as possible, I have simulated the data for each of the scales at the item level.

Simulating the data involved using factor loadings, means, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

Five additional scales were reported in the Szymanski and Bissonette article [2020]. Unfortunately, I could not locate factor loadings for all of them; and in two cases, I used estimates from a more recent psychometric analysis. When the individual item and their factor loadings are known, I assigned names based on item content (e.g., “lo\_energy”) rather than using item numbers (e.g., “PHQ4”). When I am doing psychometric analyses, I prefer item-level names so that I can quickly see (without having to look up the item names) how the items are behaving. While the focus of this series of chapters is on the LGBTQ Campus Climate scale, this simulated data might be useful to you in one or more of the suggestions for practice (e.g., examining the psychometric characteristics of one or the other scales). The scales, their original citation, and information about how I simulated data for each are listed below.

- **Sexual Orientation-Based Campus Victimization Scale** [Herek, 1993] is a 9-item item scale with Likert scaling ranging from 0 (*never*) to 3 (*two or more times*). Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.
- **College Satisfaction Scale** [Helm et al., 1998] is a 5-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores represent greater college satisfaction. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items.
- **Institutional and Goals Commitment** [Pascarella and Terenzini, 1980] is a 6-item subscale from a 35-item measure assessing academic/social integration and institutional/goal commitment (5 subscales total). The measure had with Likert scaling ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Higher scores on the institutional and goals commitment subscale indicate greater intentions to persist in college. Data were simulated using factor loadings in the source article.
- **GAD-7** [Spitzer et al., 2006] is a 7-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate more anxiety. I simulated data by estimating factor loadings from Brattmyr et al. [2022].
- **PHQ-9** [Kroenke et al., 2001] is a 9-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate higher levels of depression. I simulated data by estimating factor loadings from Brattmyr et al. [2022].

*#Entering the intercorrelations, means, and standard deviations from the journal article*

```
Szymanski_generating_model <- '
  #measurement model
  CollegeResponse  =~ .88*cold + .73*unresponsive + .73*supportive
  Stigma =~ .86*negative + .76*heterosexism + .71*harassed
  Victimization =~ .8*Vic1 + .8*Vic2 + .8*Vic3 + .8*Vic4 + .8*Vic5 + .8*Vic6 + .8*Vic7 +
  CollSat =~ .8*Sat1 + .8*Sat2 + .8*Sat3 + .8*Sat4 + .8*Sat5
  Persistence =~ .69*graduation_importance + .63*right_decision + .62*will_register + .58*will_change + .61*will_change_important
```

```

Anxiety =~ .851*nervous + .887*worry_control + .894*much_worry + 674*cant_relax + .484*
Depression =~ .798*anhedonia + .425*down + .591*sleep + .913*lo_energy + .441*appetite

#Means
CollegeResponse ~ 2.71*1
Stigma ~3.61*1
Victimization ~ 0.11*1
CollSat ~ 5.61*1
Persistence ~ 4.41*1
Anxiety ~ 1.45*1
Depression ~1.29*1

#Correlations
CollegeResponse ~~ .58*Stigma
CollegeResponse ~~ -.25*Victimization
CollegeResponse ~~ -.59*CollSat
CollegeResponse ~~ -.29*Persistence
CollegeResponse ~~ .17*Anxiety
CollegeResponse ~~ .18*Depression

Stigma ~~ .37*Victimization
Stigma ~~ -.41*CollSat
Stigma ~~ -.19*Persistence
Stigma ~~ .27*Anxiety
Stigma ~~ .24*Depression

Victimization ~~ -.22*CollSat
Victimization ~~ -.04*Persistence
Victimization ~~ .23*Anxiety
Victimization ~~ .21*Depression

CollSat ~~ .53*Persistence
CollSat ~~ -.29*Anxiety
CollSat ~~ -.32*Depression

Persistence ~~ -.22*Anxiety
Persistence ~~ -.26*Depression

Anxiety ~~ .76*Depression
'

set.seed(240218)
dfSzy <- lavaan:::simulateData(model = Szymanski_generating_model,
                                model.type = "sem",
                                meanstructure = T,
                                sample.nobs=646,

```

```

    standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(dfSzy))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 6 are the Perceptions of LGBTQ Campus Climate Scale
#Rows 7 thru 15 are the Sexual Orientation-Based Campus Victimization Scale
#Rows 16 thru 20 are the College Satisfaction Scale
#Rows 21 thru 26 are the Institutional and Goals Commitment Scale
#Rows 27 thru 33 are the GAD7
#Rows 34 thru 42 are the PHQ9

for(i in 1:ncol(dfSzy)){
  if(i >= 1 & i <= 6){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
  }
  if(i >= 7 & i <= 15){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
  }
  if(i >= 16 & i <= 20){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 7))
  }
  if(i >= 21 & i <= 26){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(1, 5))
  }
  if(i >= 27 & i <= 33){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
  }
  if(i >= 34 & i <= 42){
    dfSzy[,i] <- scales::rescale(dfSzy[,i], c(0, 3))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
dfSzy <- dfSzy %>% round(0)

#quick check of my work
#psych::describe(dfSzy)

#Reversing the supportive item on the Perceptions of LGBTQ Campus Climate Scale so that the exercises
#dfSzy <- dfSzy %>%
#  dplyr::mutate(supportiveNR = 8 - supportive)

#Reversing three items on the Institutional and Goals Commitments scale so that the exercises

```

```
dfSzy <- dfSzy %>%
  dplyr::mutate(not_graduateNR = 8 - not_graduate)%>%
  dplyr::mutate(undecidedNR = 8 - undecided)%>%
  dplyr::mutate(grades_unimportantNR = 8 - grades_unimportant)

dfSzy <- dplyr::select(dfSzy, -c(supportive, not_graduate, undecided, grades_unimportant))
```

The optional script below will let you save the simulated data to your computing environment as either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```
# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(dfSzy, 'SzyDF.rds') bring back the simulated dat from
# an .rds file dfSzy <- readRDS('SzyDF.rds')

# write the simulated data as a .csv write.table(dfSzy,
# file='SzyDF.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file dfSzy <-
# read.csv('SzyDF.csv', header = TRUE)
```

Although Szymanski and Bissonette report inter-item correlations, it does not appear that they used item analysis to guide their selection of items. In fact, it is not necessary to do so. I teach item analysis because I think it provides a conceptual grounding for future lessons on exploratory and confirmatory factor analysis.

## 7.4 Step I: Corrected item-total correlations

### Calculate corrected item-total correlations

- Reverse-score any items are negatively worded (i.e., the item is scaled opposite the other items)
- Ensure proper formatting of variables (e.g., numerical, integer)
- Evaluate *r.drop* correlations from the *psych:alpha* function in R
- Consider deleting items with low item-total correlations
  - Ask yourself, “Does this narrow the construct definition in ways that are problematic?”
  - Re-run and re-evaluate the *r.drop* analyses and alpha coefficient after each change.
  - This process is iterative and may involve “adding back” previously deleted items
  - Monitor the alpha coefficient with each addition or deletion

Figure 7.2: Image of the first step in the workflow for item analysis for survey development.

You might think of corrected item-total correlations as form a *within-scale of convergent validity*.

- If needed, transform any items (i.e., reverse-coding) and calculate a total score.
- Calculate *corrected item-total correlations* by correlating each item to the total score *excluding* the item being evaluated.
  - to the degree that the item total represents the construct of interest, the items should be strongly correlated with the corrected total score.

- Make decisions about items and scales. For items that have low or negative correlations
  - consider deletion,
  - consider revision (requires new data collection).
- Each time an item is deleted, the item analysis needs to be repeated because it changes the total-scale score.
  - In fact, it's a very iterative process. At times, researchers “add back” a previously deleted item (once others are deleted) because with each deletion/addition the statistical construct definition is evolving.
- In multidimensional scales, if the total-scale score is ever used, researchers should conduct item analyses separately for both the total- and the sub- scale scores.

There are reasons to not “blindly follow the results of an item analysis” [Green and Salkind, 2017].

- **Method factors** (aka *method effects*) are common *methods* that are irrelevant to the characteristics or traits being measured – yet when analyzed they share variance. Examples of these include negatively word items and common phrasing such as “My supervisor tells me” versus “I receive feedback” [Chyung et al., 2018c].
- Adequacy of construct representation. That is, how broad is the construct and to what degree do the items represent the entire construct? Threats to the adequacy of the construct representation include:
  - Writing items on a particular, narrow, aspect of the construct, ignoring others.
  - Retaining items that are strongly correlated while deleting those that whose correlations are less strong (although they represent a different aspect of the construct).

This means we should think carefully and simultaneously about:

- statistical properties of the item and overall scale,
- construct definition,
- scale structure (unidimensional? multidimensional? hierarchical?).

### 7.4.1 Data Prep

Let's do the operational work to get all the pieces we need:

1. Reverse-code the *supportive* variable.
2. From the raw data calculate
  - total-scale score,
  - college response subscale,
  - stigma subscale.
3. The result is dataset with the item-level data and the three mean scores (total, college response, stigma).

When we review the information about this scale, we learn that the *supportive* item is scaled in the opposite direction of the rest of the items. That is, a higher score on *supportive* would indicate a positive perception of the campus climate for LGBTQ individuals whereas higher scores on the remaining items indicate a more negative perception. Before moving forward, we must reverse score this item.

In doing this, I will briefly note that in this case I have given my variables one-word names that represent each item. Many researchers (including myself) will often give variable names that are alpha numerical: LGBTQ1, LGBTQ2, LGBTQn. Either is acceptable. In the psychometrics case, I find the the one-word names to be useful shortcuts as I begin to understand the inter-item relations.

In reverse-scoring the *supportive* item, I will rename it “*unsupportive*” as an indication of its reversed direction.

```
dfSzy <- dfSzy %>%
  dplyr::mutate(unsupportive = 8 - supportiveNR) #scaling 1 to 7; so we subtract from 8

# psych::describe(dfSzy)
```

Next, we score the items. In our simulation, we have no missing data. Using an available information approach (AIA; [Parent, 2013]) where it is common to allow 20-25% missingness, we might allow the total-scale score to calculate if there is one variable missing. I am inclined to also score the subscales if there is one missing; thus I set the threshold at 66%. The *mean\_n()* function in the *sjstats* packages is especially helpful for this.

```
LGBTQvars <- c("cold", "unresponsive", "negative", "heterosexism", "harassed",
  "unsupportive")
ResponseVars <- c("cold", "unresponsive", "unsupportive")
Stigmavars <- c("negative", "heterosexism", "harassed")

dfSzy$Total <- sjstats::mean_n(dfSzy[, LGBTQvars], 0.8) #will create the mean for each individual
dfSzy$Response <- sjstats::mean_n(dfSzy[, ResponseVars], 0.66) #will create the mean for each response
dfSzy$Stigma <- sjstats::mean_n(dfSzy[, Stigmavars], 0.66) #will create the mean for each stigma

# If the scoring code above does not work for you, try the format
# below which involves inserting periods in front of the variable
# list. One example is provided. dfLewis$Belonging <-
# sjstats::mean_n(dfLewis[, .Belonging_vars], 0.80)
```

While we are at it, let's just create tiny dfs with just our variables of interest.

```
LGBTQ <- dplyr::select(dfSzy, cold, unresponsive, unsupportive, negative,
  heterosexism, harassed)
Response <- dplyr::select(dfSzy, cold, unresponsive, unsupportive)
Stigma <- dplyr::select(dfSzy, negative, heterosexism, harassed)
```

### 7.4.2 Calculating Item-Total Correlation Coefficients

Let's first ask, "Is there support for this instrument as a unidimensional measure?" To do that, we get an alpha for the whole scale score.

The easiest way to do this is apply the *alpha()* function to a tiny df with the variables in that particular scale or subscale. Any variables should be pre-reversed.

```
LGBTQalpha <- psych::alpha(LGBTQ) #Although unnecessary, I have saved the output as objects b
LGBTQalpha
```

```
Reliability analysis
Call: psych::alpha(x = LGBTQ)
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.7	0.7	0.68	0.28	2.4	0.018	4	0.63	0.25

```
95% confidence boundaries
      lower alpha upper
Feldt     0.66   0.7  0.74
Duhachek 0.66   0.7  0.74
```

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
cold	0.64	0.64	0.61	0.27	1.8	0.022	0.0066	0.22	
unresponsive	0.66	0.66	0.63	0.28	2.0	0.021	0.0073	0.25	
unsupportedive	0.67	0.67	0.63	0.29	2.0	0.021	0.0058	0.25	
negative	0.66	0.66	0.63	0.28	2.0	0.021	0.0084	0.25	
heterosexism	0.66	0.66	0.63	0.28	2.0	0.021	0.0087	0.25	
harassed	0.67	0.67	0.64	0.29	2.0	0.021	0.0078	0.25	

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
cold	646	0.68	0.68	0.59	0.49	4.1	1.03
unresponsive	646	0.63	0.63	0.51	0.43	4.3	0.99
unsupportedive	646	0.62	0.62	0.51	0.42	3.7	0.98
negative	646	0.64	0.63	0.51	0.42	4.0	1.04
heterosexism	646	0.61	0.63	0.51	0.43	4.0	0.90
harassed	646	0.63	0.61	0.49	0.41	3.9	1.07

Non missing response frequency for each item

	1	2	3	4	5	6	7	miss
cold	0.00	0.04	0.22	0.40	0.23	0.09	0.00	0
unresponsive	0.00	0.03	0.17	0.37	0.33	0.09	0.01	0
unsupportedive	0.01	0.07	0.35	0.37	0.17	0.02	0.01	0
negative	0.01	0.07	0.23	0.39	0.24	0.05	0.00	0
heterosexism	0.00	0.03	0.24	0.43	0.26	0.03	0.00	0

```
harassed    0.01 0.07 0.27 0.37 0.22 0.05 0.01      0
```

Examining our list, the overall alpha is 0.70. Further, the average inter-item correlation (*average\_r*) is .28.

*And just hold up a minute, I thought you told us alpha was bad!*

- While it is less than ideal, we still use it all the time:
  - keeping in mind its relative value (does it increase/decrease, holding other things [like sample size] constant) and
  - examining alpha alternatives (such as we obtained from the omega output)
- Why alpha in this context? Its information about *consistency* is essential. In evaluating a scale's reliability we do want to know if items (unidimensionally or across subscales) are responding consistently high/middle/low.

We take note of two columns:

- *r.cor* is the correlation between the item and the total-scale score with the row-item included. When our focus is on the contribution of a specific item, this information is not helpful since this column gets “extra credit” for the redundancy of the duplicated item.
- *r.drop* is the corrected item-total correlation. This is the better choice because it excludes the row-item being evaluated (eliminating the redundancy) prior to conducting the correlation.
  - Looking at the two columns, notice that the *r.drop* correlations are lower. This is the more honest correlation of the item with the *other* items.
  - In item analysis, we look for items that have relatively high (assessing redundancy or duplication) or items and relatively low (indicating they are unlike the other items) values.

If we thought an item was problematic, we could eliminate it and rerun the analysis. Because we are looking at a list of items that “made the cut,” we don’t have any items that are concerningly high or low. For demonstration purposes, though, the corrected item-total correlation (*r.drop*) of the *harassed* variable was the lowest (0.40). Let’s re-run the analysis excluding this item.

```
minus_harassed <- dplyr::select(dfSzy, cold, unresponsive, unsupportive,
                                 negative, heterosexism)
```

```
psych::alpha(minus_harassed)
```

```
Reliability analysis
Call: psych::alpha(x = minus_harassed)
```

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.67	0.67	0.64	0.29	2	0.021	4	0.65	0.25

```

95% confidence boundaries
      lower alpha upper
Feldt      0.63  0.67  0.71
Duhachek   0.63  0.67  0.71

Reliability if an item is dropped:
      raw_alpha std.alpha G6(smc) average_r S/N alpha se  var.r med.r
cold          0.58      0.58    0.53     0.26 1.4   0.027 0.0060  0.22
unresponsive   0.62      0.62    0.56     0.29 1.6   0.025 0.0079  0.25
unsupportive   0.61      0.61    0.56     0.28 1.6   0.025 0.0072  0.25
negative       0.65      0.64    0.58     0.31 1.8   0.022 0.0093  0.30
heterosexism   0.64      0.64    0.58     0.31 1.8   0.023 0.0097  0.30

Item statistics
      n raw.r std.r r.cor r.drop mean   sd
cold      646  0.72  0.71  0.61   0.50  4.1 1.03
unresponsive 646  0.66  0.66  0.53   0.43  4.3 0.99
unsupportive 646  0.67  0.67  0.55   0.45  3.7 0.98
negative     646  0.63  0.62  0.46   0.37  4.0 1.04
heterosexism 646  0.60  0.62  0.47   0.38  4.0 0.90

Non missing response frequency for each item
      1    2    3    4    5    6    7 miss
cold    0.00 0.04 0.22 0.40 0.23 0.09 0.00   0
unresponsive 0.00 0.03 0.17 0.37 0.33 0.09 0.01   0
unsupportive 0.01 0.07 0.35 0.37 0.17 0.02 0.01   0
negative     0.01 0.07 0.23 0.39 0.24 0.05 0.00   0
heterosexism 0.00 0.03 0.24 0.43 0.26 0.03 0.00   0

```

The alpha ( $\alpha = 0.67$ ) decreases; the overall inter-item correlations increase, slightly (*average\_r*; 0.29). This decrease in alpha is an example of how sample size can effect the result.

Examining item-level statistics, we do see greater variability (0.37 to 0.50) in the corrected item-total correlations (*r.drop*). What might this mean?

- The item we dropped (*harassed*) may be clustering with *negative* and *heterosexism* in a subordinate factor (think subscale).
- Although item analysis is more of a tool in assessing reliability, the statistical information that *harassed* provided may broaden the construct definition (definitions are a concern of *validity*) of perceptions of campus climate such that it is necessary to ground/anchor *negative* and *heterosexism*.

Tentative conclusion: there is evidence that this is not a unidimensional measure. Let's move on to inspect similar data for each of the subscales. We'll start with the College Response subscale.

```

RESPAlpha <- psych::alpha(Response)
RESPAlpha

```

Reliability analysis

Call: psych::alpha(x = Response)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r	
0.66	0.66	0.57		0.39	1.9	0.023	4	0.77	0.4

95% confidence boundaries

	lower	alpha	upper
Feldt	0.61	0.66	0.70
Duhachek	0.62	0.66	0.71

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
cold	0.52	0.52	0.35		0.35	1.1	0.038	NA	0.35
unresponsive	0.60	0.60	0.42		0.42	1.5	0.032	NA	0.42
unsupportive	0.58	0.58	0.40		0.40	1.4	0.033	NA	0.40

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
cold	646	0.80	0.79	0.62	0.50	4.1	1.03
unresponsive	646	0.76	0.76	0.55	0.45	4.3	0.99
unsupportive	646	0.76	0.77	0.57	0.46	3.7	0.98

Non missing response frequency for each item

	1	2	3	4	5	6	7	miss
cold	0.00	0.04	0.22	0.40	0.23	0.09	0.00	0
unresponsive	0.00	0.03	0.17	0.37	0.33	0.09	0.01	0
unsupportive	0.01	0.07	0.35	0.37	0.17	0.02	0.01	0

The alpha for the College Response subscale is 0.66; this is a bit lower than the alpha for the total scale score ( $\alpha = 0.70$ ). The average inter-item correlation (*average\_r*) is higher somewhat higher than the that of the total scale score (0.39 versus 0.28).

Examining the corrected item-total correlations (*r.drop*) indicates strong correlations between the row-item with the remaining variables (0.45 to 0.50).

Let's examine at the Stigma subscale.

```
STIGalpha <- psych::alpha(Stigma)
STIGalpha
```

Reliability analysis

Call: psych::alpha(x = Stigma)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r	
0.62	0.63	0.53		0.36	1.7	0.025	4	0.76	0.36

```

95% confidence boundaries
    lower alpha upper
Feldt      0.57  0.62  0.67
Duhachek   0.57  0.62  0.67

Reliability if an item is dropped:
    raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
negative       0.52      0.52    0.35      0.35 1.1    0.038    NA  0.35
heterosexism   0.53      0.53    0.36      0.36 1.1    0.037    NA  0.36
harassed       0.53      0.54    0.37      0.37 1.2    0.036    NA  0.37

Item statistics
    n raw.r std.r r.cor r.drop mean   sd
negative     646  0.77  0.76  0.56   0.44  4.0 1.0
heterosexism 646  0.73  0.76  0.55   0.44  4.0 0.9
harassed     646  0.77  0.75  0.54   0.43  3.9 1.1

Non missing response frequency for each item
    1   2   3   4   5   6   7 miss
negative     0.01 0.07 0.23 0.39 0.24 0.05 0.00   0
heterosexism 0.00 0.03 0.24 0.43 0.26 0.03 0.00   0
harassed     0.01 0.07 0.27 0.37 0.22 0.05 0.01   0

```

The alpha for the Stigma subscale is 0.63; this is a bit lower than the alpha for the total-scale ( $\alpha = 0.70$ ). In contrast, the inter-item correlation (*average\_r*) is a bit higher than the same for the total scale score (0.36 versus 0.28).

Examining the corrected item-total correlations (*r.drop*) indicates a strong correlation between the row-item with the remaining variables (0.43 to 0.44).

In addition to needing strong inter-item correlations (which we just assessed) we want the individual items to correlate more strongly with themselves than with the other scale. Let's do that next.

## 7.5 Step II: Correlating Items with Other Scale Totals

You might think of this step as analyzing a within-scale version of discriminant validity. That is, we do not want individual items from one scale to correlate more highly with subscale scores of other scales, than it does with its own.

### Calculate correlations of items with other subscale scores

- Calculate the mean scores for each of the subscales of a measure.
- Focusing on one subscale at a time, correlate each of the subscale's items with the total scores of all the other subscales

Figure 7.3: Image of the second step in the workflow for item analysis for survey development.

- Calculate scale scores for each of the subscales of a measure.

- Focusing on one subscale at a time, correlate each of the subscale's items with the total scores of all the other subscales.
- Comparing to the results of Step I's corrected item-total process, each item should have stronger correlations with its own items (i.e., the corrected item-total correlations) than with the other subscale total scores.

In this first analysis, we will correlate the individual *items* from the College Response subscale to the Stigma subscale *score*.

```
apaTables::apa.cor.table(dfSzy[c("cold", "unresponsive", "unsupportive",
  "Stigma")])
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. cold	4.11	1.03			
2. unresponsive	4.31	0.99	.40** [.34, .47]		
3. unsupportive	3.69	0.98	.42** [.36, .49]	.35** [.28, .42]	
4. Stigma	3.96	0.76	.33** [.26, .39]	.28** [.21, .35]	.26** [.18, .33]

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

We want the corrected item-total correlations of the College Response scale (0.45 to 0.50; retrieved from the *r.drop* column above) to be higher than their correlations with the Stigma scale (0.28 to 0.33) with all three items). These items follow the pattern.

Let's examine the individual items from the Stigma scale with the College Response subscale score.

```
apaTables::apa.cor.table(dfSzy[c("negative", "heterosexism", "harassed",
  "Response")])
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. negative	3.96	1.04			
2. heterosexism	4.00	0.90	.37**		
			[.30, .43]		
3. harassed	3.91	1.07	.36**	.35**	
			[.29, .42]	[.28, .42]	
4. Response	4.04	0.77	.29**	.29**	.27**
			[.22, .36]	[.22, .36]	[.20, .34]

Note. M and SD are used to represent mean and standard deviation, respectively.

Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

Similarly, the corrected item-total correlations (i.e.,  $r.drop$ ) from the Stigma subscale (0.43 to 0.44) are stronger than their correlation with the College Response subscale (0.27 to 0.29).

## 7.6 Step III: Interpreting and Writing up the Results

Now it's time to make sense of the results. Here's a reminder from the workflow:

- Compare the corrected item-total correlations to the correlations of items with other subscale scores**
  - Corrected item-total correlations should be stronger/higher than the correlations of items with other scales' subscores

Figure 7.4: Image of the third step in the workflow for item analysis for survey development.

Tabling these results can be really useful to present them effectively. As is customary in APA style tables, when the item is in bold, the value represents its relationship with its own factor. These values come from the corrected item-total ( $r.drop$ ) values where the item is singled out and correlated with the remaining items in its subscale.

---

Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	College Response	Stigma
cold	.50	0.33
unresponsive	.45	0.28
unsupportive	.46	0.26
negative	0.29	<b>0.44</b>
heterosexism	0.29	<b>0.44</b>
harassed	0.27	<b>0.43</b>

Although I pitched this type of item-analysis as *reliability*, to some degree it assesses within-scale **convergent and discriminant validity** because we can see the item relates more strongly to members of its own scale (higher correlation coefficients indicate *convergence*) than to the subscale scores of the other scales. When this pattern occurs, we can argue that the items *discriminate* well.

## Results

Item analyses were conducted on the six items hypothesized to assess perceptions of campus climate for members of the LGBTQ community. To assess the within-scale convergent and discriminant validity of the College Response and Stigma subscales, each item was correlated with its own scale (with the item removed) and with the other subscale (see Table 1). In all cases, items were more highly correlated with their own scale than with the other scale. Coefficient alphas were 0.66, 0.63, and 0.70 for the College Response, Stigma, and total-scale scores, respectively. We concluded that the within-scale convergent and discriminant validity of this measure is strong.

For your consideration: You are at your dissertation defense. For one of your measures, the Cronbach's alpha is .45. A committee member asks, "So why was the alpha coefficient so low?" On the basis of what you have learned in this lesson, how do you respond?

## 7.7 A Conversation with Dr. Szymanski

Doctoral students Julian Williams (Industrial-Organizational Psychology), Jaylee York (Clinical Psychology), and I were able to interview the first author (Dawn Szymanski, PhD) about the article [[Szymanski and Bissonette, 2020](#)] and what it means. Here's a direct [link](#) to that interview.

Among other things, we asked:

- How were you able to create such an efficient (6 items) survey?
- What were the decisions around a potential third factor of *visibility*?
- What would you say to senior leadership on a college campus (where there hiring policies that discriminate against LGBTQIA+ applicants) who will acknowledge the research that indicates that the existence of such policies are associated with reduced well-being for members of the LGBTQIA+ community but who insists that their campus is different?
- How would you like to see the article used?

## 7.8 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. For this lesson, please locate item-level data for a scale that

has the potential for at least two subscales and a total-scale score. Ideally, you selected such data for practice from the prior lesson. Then, please examine the following:

- produce alpha coefficients, average inter-item correlations, and corrected item-total correlations for the total and subscales, separately
- produce correlations between the individual items of one subscale and the subscale scores of all other scales
- draft an APA style results section with an accompanying table.

In my example, there were only two subscales. If you have more, you will need to compare each subscale with all the others. For example, if you had three subscales: A, B, C, you would need to compare A/B, B/C, and A/C.

### 7.8.1 Problem #1: Play around with this simulation.

Copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results.

If item analysis is new to you, copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results. Perhaps you just change the number in “set.seed(210827)” from 210827 to something else. Your results should parallel those obtained in the lecture, making it easier for you to check your work as you go.

### 7.8.2 Problem #2: Use raw data from the ReCentering Psych Stats survey on Qualtrics.

The script below pulls live data directly from the ReCentering Psych Stats survey on Qualtrics. As described in the [Scrubbing and Scoring chapters](#) of the ReCentering Psych Stats Multivariate Modeling volume, the Perceptions of the LGBTQ College Campus Climate Scale [Szymanski and Bissonette, 2020] was included (LGBTQ) and further adapted to assess perceptions of campus climate for Black students (BLst), non-Black students of color (nBSoc), international students (INTst), and students with disabilities (wDIS). Consider conducting the analyses on one of these scales or merging them together and imagining subscales according to identity/group (LGBTQ, Black, non-Black, disability, international) or College Response and Stigma across the different groups.

```
library(tidyverse)
# only have to run this ONCE to draw from the same Qualtrics
# account...but will need to get different token if you are changing
# between accounts
library(qualtRics)
# qualtrics_api_credentials(api_key =
#   'mUgPMMySYkiWpMFkwHale1QE5HNmh5LRUaA8d9PDg', base_url =
#   'spupsych.az1.qualtrics.com', overwrite = TRUE, install = TRUE)
QTRX_df <- qualtRics::fetch_survey(surveyID = "SV_b2cClqA1LGQ6nLU", time_zone = NULL,
  verbose = FALSE, label = FALSE, convert = FALSE, force_request = TRUE,
```

```

import_id = FALSE)
climate_df <- QTRX_df %>%
  select("Blst_1", "Blst_2", "Blst_3", "Blst_4", "Blst_5", "Blst_6",
         "nBSoc_1", "nBSoc_2", "nBSoc_3", "nBSoc_4", "nBSoc_5", "nBSoc_6",
         "INTst_1", "INTst_2", "INTst_3", "INTst_4", "INTst_5", "INTst_6",
         "wDIS_1", "wDIS_2", "wDIS_3", "wDIS_4", "wDIS_5", "wDIS_6", "LGBTQ_1",
         "LGBTQ_2", "LGBTQ_3", "LGBTQ_4", "LGBTQ_5", "LGBTQ_6")
# Item numbers are supported with the following items: _1 'My campus
# unit provides a supportive environment for ___ students' _2
# '_____ is visible in my campus unit' _3 'Negative attitudes
# toward persons who are ___ are openly expressed in my campus
# unit.' _4 'My campus unit is unresponsive to the needs of ___
# students.' _5 'Students who are___ are harassed in my campus
# unit.' _6 'My campus unit is cold and uncaring toward ___
# students.'

# Item 1 on each subscale should be reverse coded. The College
# Response scale is composed of items 1, 4, 6, The Stigma scale is
# composed of items 2,3, 5

```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do).

```

# write the simulated data as a .csv write.table(climate_df,
# file='climate_df.csv', sep=',', col.names=TRUE, row.names=FALSE)
# bring back the simulated dat from a .csv file climate_df <-
# read.csv ('climate_df.csv', header = TRUE)

```

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(climate_df, 'climate_df.rds') bring back the simulated
# dat from an .rds file climate_df <- readRDS('climate_df.rds')

```

### 7.8.3 Problem #3: Try something entirely new.

Complete the same steps using data for which you have permission and access. This might be data of your own, from your lab, simulated from an article, or located on an open repository.

### 7.8.4 Grading Rubric

Using the lecture and workflow (chart) as a guide, please work through all the steps listed in the proposed assignment/grading rubric.

Assignment Component	Points Possible	Points Earned
1. Check and, if needed, format and score data	5	_____
2. Report alpha coefficients and average inter-item correlations for the total and subscales	5	_____
3. Produce and interpret corrected item-total correlations for total and subscales, separately	5	_____
4. Produce and interpret correlations between the individual items of a given subscale and the subscale scores of all other subscales	5	_____
5. APA style results section with table	5	_____
6. Explanation to grader	5	_____
<b>Totals</b>	<b>30</b>	_____

## 7.9 Bonus Reel:

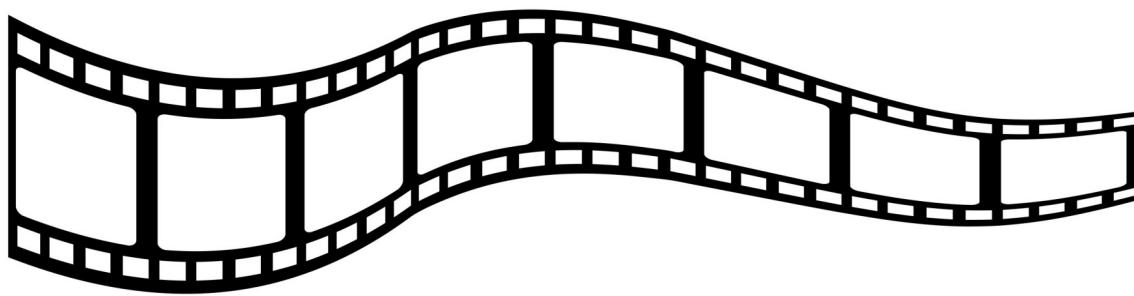


Figure 7.5: Image of a filmstrip

For our interpretation and results, I created the table by manually typing in the results. Since there were only two subscales, this was easy. However, it can be a very useful skill (and prevent typing errors) by leveraging R's capabilities to build a table.

The script below

- Creates a correlation matrix of the items of each scale and correlates them with the “other” subscale, separately for both subscales.
- Extracts the r.drop from each subscale
- Joins (adds more variables) the analyses across the corrected item-total and item-other subscale analyses
- Binds (adds more cases) the two sets of items together

```
Resp_othR <- psych::corr.test(dfSzy[c("negative", "heterosexism", "harassed",
  "Response")]) #Run the correlation of the subscale and the items that are *not* on the su
Resp_othR <- as.data.frame(Resp_othR$r) #extracts the 'r' matrix and makes it a df
```

```

Resp_othR$Items <- c("negative", "heterosexism", "harassed", "Response") #Assigning names to
Resp_othR <- Resp_othR[!Resp_othR$Items == "Response", ] #Removing the subscale score as a a
Resp_othR[, "Stigma"] <- NA #We need a column for this to bind the items later
Resp_othR <- dplyr::select(Resp_othR, Items, Response, Stigma) #All we need is the item name
RESPAlpha <- as.data.frame(RESPAlpha$item.stats) #Grabbing the alpha objet we created earlier
RESPAlpha$Items <- c("cold", "unresponsive", "unsupportive")

Stig_othR <- psych::corr.test(dfSzy[c("cold", "unresponsive", "unsupportive",
    "Stigma")]) #Run the correlation of the subscale and the items that are *not* on the subscale
Stig_othR <- as.data.frame(Stig_othR$r) #extracts the 'r' matrix and makes it a df
Stig_othR$Items <- c("cold", "unresponsive", "unsupportive", "Stigma") #Assigning names to the items
Stig_othR <- Stig_othR[!Stig_othR$Items == "Stigma", ] #Removing the subscale score as a a row
Stig_othR[, "Response"] <- NA #We need a column for this to bind the items later
Stig_othR <- dplyr::select(Stig_othR, Items, Response, Stigma) #All we need is the item name
STIGAlpha <- as.data.frame(STIGAlpha$item.stats) #Grabbing the alpha objet we created earlier
STIGAlpha$Items <- c("negative", "heterosexism", "harassed")

# Combining these four dfs
ResponseStats <- full_join(RESPAlpha, Stig_othR, by = "Items")
ResponseStats$Response <- ResponseStats$r.drop
ResponseStats <- dplyr::select(ResponseStats, Items, Response, Stigma)
StigmaStats <- full_join(STIGAlpha, Resp_othR, by = "Items")
StigmaStats$Stigma <- StigmaStats$r.drop
StigmaStats <- dplyr::select(StigmaStats, Items, Response, Stigma)
ItemAnalyses <- rbind(ResponseStats, StigmaStats)
ItemAnalyses

```

	Items	Response	Stigma
1	cold	0.5041391	0.3258401
2	unresponsive	0.4490390	0.2783617
3	unsupportive	0.4644473	0.2581391
4	negative	0.2874243	0.4401670
5	heterosexism	0.2925477	0.4365623
6	harassed	0.2704033	0.4291150

```

# Writing them to a .csv file allows post-r formatting
write.csv(ItemAnalyses, file = "LGBTQ_Climate_ItemAnalyses.csv", sep = ",",
          row.names = TRUE, col.names = TRUE)

```

## 7.10 Homeworked Example

### Screencast Link

For more information about the data used in this homeworked example, please refer to the description and codebook located at the end of the [introduction](#) in first volume of ReCentering Psych Stats.

As a brief review, this data is part of an IRB-approved study, with consent to use in teaching demonstrations and to be made available to the general public via the open science framework. Hence, it is appropriate to use in this context. You will notice there are student- and teacher- IDs. These numbers are not actual student and teacher IDs, rather they were further re-identified so that they could not be connected to actual people.

Because this is an actual dataset, if you wish to work the problem along with me, you will need to download the [ReC.rds](#) data file from the Worked\_Examples folder in the ReC\_Psychometrics project on the GitHub.

The course evaluation items can be divided into three subscales:

- **Valued by the student** includes the items: ValObjectives, IncrUnderstanding, IncrInterest
- **Traditional pedagogy** includes the items: ClearResponsibilities, EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation
- **Socially responsive pedagogy** includes the items: InclusvClassrm, EquitableEval, MultiPerspectives, DEIIntegration

In this homework focused on validity we will score the total scale and subscales, create a correlation matrix of our scales with a different scale (or item), formally test to see if correlation coefficients are statistically significantly different from each other, conduct a test of incremental validity.

### 7.10.1 Check and, if needed, format and score data

```
big <- readRDS("ReC.rds")
```

With the next code I will create an item-level df with only the items used in the three scales.

```
library(tidyverse)
items <- big %>%
  dplyr::select (ValObjectives, IncrUnderstanding, IncrInterest, ClearResponsibilities, EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation)
```

Next I check the structure of the data.

```
str(items)
```

```
Classes 'data.table' and 'data.frame': 310 obs. of 12 variables:
 $ ValObjectives      : int  5 5 4 4 5 5 5 5 4 5 ...
 $ IncrUnderstanding   : int  2 3 4 3 4 4 5 2 4 5 ...
 $ IncrInterest        : int  5 3 4 2 4 3 5 3 2 5 ...
 $ ClearResponsibilities: int  5 5 4 4 5 4 5 4 4 5 ...
 $ EffectiveAnswers    : int  5 3 5 3 5 3 4 3 2 3 ...
 $ Feedback             : int  5 3 4 2 5 NA 5 4 4 5 ...
 $ ClearOrganization    : int  3 4 3 4 4 4 5 4 4 5 ...
 $ ClearPresentation    : int  4 4 4 2 5 3 4 4 4 5 ...
```

```
$ MultPerspectives      : int  5 5 4 5 5 4 5 5 5 5 ...
$ InclusvClassrm       : int  5 5 5 5 5 4 5 5 4 5 ...
$ DEIintegration        : int  5 5 5 5 5 4 5 5 5 5 ...
$ EquitableEval         : int  5 5 3 5 5 3 5 5 3 5 ...
- attr(*, ".internal.selfref")=<externalptr>
```

### 7.10.2 Report alpha coefficients and average inter-item correlations for the total and subscales

This task is completed in the next section.

### 7.10.3 Produce and interpret corrected item-total correlations for total and subscales, separately

In the lecture, I created baby dfs of the subscales and ran the alphas on those; another option is to create lists of variables (i.e., variable vectors) and use that instead. We can later use those same variable vectors to score the items.

#### 7.10.3.1 All Course Evaluation Items

First, I will calculate the alpha coefficient and inter-item correlations for the total score representing the course evaluation items.

```
psych::alpha(items)
```

```
Reliability analysis
Call: psych::alpha(x = items)
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.92	0.92	0.93	0.49	11	0.0065	4.3	0.61	0.48

```
95% confidence boundaries
      lower alpha upper
Feldt      0.90  0.92  0.93
Duhachek   0.91  0.92  0.93
```

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
ValObjectives	0.92	0.92	0.93	0.51	11.3	0.0067	0.016	
IncrUnderstanding	0.91	0.91	0.92	0.49	10.6	0.0070	0.016	
IncrInterest	0.91	0.91	0.92	0.49	10.4	0.0070	0.018	
ClearResponsibilities	0.91	0.91	0.92	0.48	10.0	0.0073	0.015	
EffectiveAnswers	0.91	0.91	0.92	0.48	10.0	0.0074	0.016	
Feedback	0.91	0.91	0.92	0.48	10.3	0.0071	0.018	

ClearOrganization	0.91	0.91	0.92	0.48	10.2	0.0073	0.016
ClearPresentation	0.91	0.91	0.92	0.47	9.7	0.0076	0.015
MultPerspectives	0.91	0.91	0.92	0.48	10.0	0.0073	0.017
InclusvClassrm	0.91	0.91	0.92	0.49	10.6	0.0069	0.018
DEIIntegration	0.92	0.92	0.93	0.52	11.8	0.0063	0.011
EquitableEval	0.91	0.91	0.93	0.49	10.5	0.0070	0.018
med.r							
ValObjectives	0.53						
IncrUnderstanding	0.50						
IncrInterest	0.48						
ClearResponsibilities	0.48						
EffectiveAnswers	0.48						
Feedback	0.48						
ClearOrganization	0.48						
ClearPresentation	0.47						
MultPerspectives	0.47						
InclusvClassrm	0.52						
DEIIntegration	0.53						
EquitableEval	0.48						

## Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
ValObjectives	309	0.59	0.61	0.55	0.53	4.5	0.61
IncrUnderstanding	309	0.71	0.70	0.67	0.64	4.3	0.82
IncrInterest	308	0.75	0.73	0.71	0.68	3.9	0.99
ClearResponsibilities	307	0.80	0.80	0.79	0.75	4.4	0.82
EffectiveAnswers	308	0.80	0.79	0.78	0.75	4.4	0.83
Feedback	304	0.75	0.75	0.72	0.69	4.2	0.88
ClearOrganization	309	0.79	0.77	0.75	0.72	4.0	1.08
ClearPresentation	309	0.85	0.84	0.83	0.80	4.2	0.92
MultPerspectives	305	0.79	0.80	0.78	0.75	4.4	0.84
InclusvClassrm	301	0.68	0.70	0.67	0.62	4.6	0.68
DEIIntegration	273	0.51	0.53	0.49	0.42	4.5	0.74
EquitableEval	308	0.70	0.72	0.69	0.66	4.6	0.63

## Non missing response frequency for each item

	1	2	3	4	5	miss
ValObjectives	0.00	0.01	0.03	0.39	0.57	0.00
IncrUnderstanding	0.01	0.04	0.07	0.44	0.45	0.00
IncrInterest	0.02	0.09	0.14	0.44	0.31	0.01
ClearResponsibilities	0.01	0.02	0.07	0.31	0.59	0.01
EffectiveAnswers	0.01	0.02	0.08	0.36	0.53	0.01
Feedback	0.01	0.05	0.10	0.39	0.46	0.02
ClearOrganization	0.04	0.07	0.10	0.41	0.38	0.00
ClearPresentation	0.02	0.05	0.07	0.40	0.46	0.00
MultPerspectives	0.02	0.02	0.08	0.33	0.56	0.02
InclusvClassrm	0.01	0.01	0.05	0.23	0.70	0.03
DEIIntegration	0.00	0.01	0.10	0.22	0.67	0.12

```
EquitableEval      0.00 0.01 0.03 0.32 0.63 0.01
```

At the total scale level,  $\alpha = 0.92$ ; the average inter-item correlations are 0.487; and the corrected item-total correlations ( $r.drop$ ) range from 0.42 to 0.80.

To obtain this data at the subscale level I will first create variable vectors.

```
Valued_vars <- c('ValObjectives', 'IncrUnderstanding', 'IncrInterest')
TradPed_vars <- c('ClearResponsibilities', 'EffectiveAnswers', 'Feedback', 'ClearOrganization')
SCRPed_vars <- c('MultPerspectives', 'InclusvClassrm', 'DEIintegration', 'EquitableEval')
```

I can insert these variable vectors into the *psych::alpha()* function to obtain the information.

### 7.10.3.2 Valued-by-Student Subscale

```
psych::alpha(items[, Valued_vars])
```

Reliability analysis

Call: psych::alpha(x = items[, Valued\_vars])

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.77	0.77	0.71	0.53	3.4	0.02	4.2	0.68	0.48

95% confidence boundaries

	lower	alpha	upper
Feldt	0.72	0.77	0.81
Duhachek	0.73	0.77	0.81

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
ValObjectives	0.80	0.81	0.68	0.68	4.3	0.022	NA	
IncrUnderstanding	0.60	0.65	0.48	0.48	1.8	0.040	NA	
IncrInterest	0.59	0.61	0.44	0.44	1.6	0.044	NA	
								med.r
ValObjectives		0.68						
IncrUnderstanding		0.48						
IncrInterest		0.44						

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
ValObjectives	309	0.71	0.77	0.55	0.50	4.5	0.61
IncrUnderstanding	309	0.86	0.85	0.76	0.68	4.3	0.82
IncrInterest	308	0.90	0.87	0.79	0.70	3.9	0.99

Non missing response frequency for each item

	1	2	3	4	5	miss
ValObjectives	0.00	0.01	0.03	0.39	0.57	0.00
IncrUnderstanding	0.01	0.04	0.07	0.44	0.45	0.00
IncrInterest	0.02	0.09	0.14	0.44	0.31	0.01

The alpha for Valued-by-the-Student subscale is 0.77. I will start building my homework table as I go along by adding the corrected item-total correlations (i.e., the *r.drop* column).

---

Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50		
IncrUnderstanding	.68		
IncrInterest	.70		
ClearResponsibilities			
EffectiveAnswers			
Feedback			
ClearOrganization			
ClearPresentation			
MultPerspectives			
DEIntegration			
EquitableEval			
Feedback			

---

### 7.10.3.3 Traditional Pedagogy Items

Next I will produce the output for the Traditional Pedagogy subscale.

```
psych::alpha(items[,TradPed_vars])
```

```
Reliability analysis
```

```
Call: psych::alpha(x = items[, TradPed_vars])
```

```
raw_alpha std.alpha G6(smc) average_r S/N      ase mean    sd median_r
 0.89        0.9       0.88       0.64 8.8 0.0094   4.3 0.76       0.65
```

```
95% confidence boundaries
```

```
lower alpha upper
```

```
Feldt     0.87  0.89  0.91
```

```
Duhachek 0.88  0.89  0.91
```

```
Reliability if an item is dropped:
```

```
raw_alpha std.alpha G6(smc) average_r S/N alpha se  var.r
```

ClearResponsibilities	0.86	0.86	0.84	0.62	6.4	0.013	0.0054
EffectiveAnswers	0.87	0.87	0.84	0.63	6.8	0.012	0.0045
Feedback	0.89	0.89	0.87	0.68	8.4	0.010	0.0016
ClearOrganization	0.88	0.88	0.85	0.64	7.2	0.012	0.0044
ClearPresentation	0.86	0.87	0.83	0.62	6.5	0.013	0.0030

med.r

ClearResponsibilities	0.59
EffectiveAnswers	0.65
Feedback	0.69
ClearOrganization	0.66
ClearPresentation	0.62

#### Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
ClearResponsibilities	307	0.87	0.87	0.84	0.79	4.4	0.82
EffectiveAnswers	308	0.84	0.85	0.81	0.76	4.4	0.83
Feedback	304	0.78	0.79	0.70	0.66	4.2	0.88
ClearOrganization	309	0.85	0.83	0.78	0.74	4.0	1.08
ClearPresentation	309	0.87	0.87	0.83	0.78	4.2	0.92

#### Non missing response frequency for each item

	1	2	3	4	5	miss
ClearResponsibilities	0.01	0.02	0.07	0.31	0.59	0.01
EffectiveAnswers	0.01	0.02	0.08	0.36	0.53	0.01
Feedback	0.01	0.05	0.10	0.39	0.46	0.02
ClearOrganization	0.04	0.07	0.10	0.41	0.38	0.00
ClearPresentation	0.02	0.05	0.07	0.40	0.46	0.00

Traditional Pedagogy  $\alpha = 0.90$ . I retrieve the corrected item-total correlations from the *r.drop* column.

---

#### Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50		
IncrUnderstanding	.67		
IncrInterest	.70		
ClearResponsibilities		.79	
EffectiveAnswers		.76	
Feedback		.66	
ClearOrganization		.74	
ClearPresentation		.78	
MultPerspectives			
DEIntegration			
EquitableEval			

Variables	Valued	TradPed	SCRPed
Feedback			

#### 7.10.3.4 Socially Responsive Pedagogy Items

```
psych::alpha(items[, SCRPed_vars])
```

Reliability analysis

Call: psych::alpha(x = items[, SCRPed\_vars])

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.81	0.81	0.78	0.52	4.3	0.017	4.5	0.58	0.54

95% confidence boundaries

lower alpha upper

Feldt	0.77	0.81	0.84
-------	------	------	------

Duhachek	0.77	0.81	0.84
----------	------	------	------

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
MultPerspectives	0.73	0.74	0.67	0.49	2.8	0.026	0.0153	
InclusvClassrm	0.74	0.74	0.67	0.49	2.9	0.025	0.0120	
DEIIntegration	0.78	0.78	0.71	0.54	3.6	0.021	0.0044	
EquitableEval	0.78	0.79	0.73	0.56	3.9	0.021	0.0034	

med.r

MultPerspectives	0.47
------------------	------

InclusvClassrm	0.50
----------------	------

DEIIntegration	0.57
----------------	------

EquitableEval	0.57
---------------	------

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
MultPerspectives	305	0.85	0.83	0.76	0.68	4.4	0.84
InclusvClassrm	301	0.82	0.83	0.76	0.67	4.6	0.68
DEIIntegration	273	0.78	0.78	0.67	0.59	4.5	0.74
EquitableEval	308	0.75	0.76	0.64	0.58	4.6	0.63

Non missing response frequency for each item

1	2	3	4	5	miss
---	---	---	---	---	------

MultPerspectives	0.02	0.02	0.08	0.33	0.56	0.02
------------------	------	------	------	------	------	------

InclusvClassrm	0.01	0.01	0.05	0.23	0.70	0.03
----------------	------	------	------	------	------	------

DEIIntegration	0.00	0.01	0.10	0.22	0.67	0.12
----------------	------	------	------	------	------	------

EquitableEval	0.00	0.01	0.03	0.32	0.63	0.01
---------------	------	------	------	------	------	------

Alpha for the SCR Pedagogy dimension is .81

Item-Total Correlations of Items with their Own and Other Subscales

Variables	Valued	TradPed	SCRPed
ValObjectives	.50		
IncrUnderstanding	.67		
IncrInterest	.70		
ClearResponsibilities		.79	
EffectiveAnswers		.76	
Feedback		.66	
ClearOrganization		.74	
ClearPresentation		.78	
MultPerspectives			.68
DEIntegration			.67
EquitableEval			.59
Feedback			.58

#### 7.10.4 Produce and interpret correlations between the individual items of a given subscale and the subscale scores of all other subscales

To do this we need to have subscale scores. Conveniently, I can use the variable vectors created in an earlier step. I will score each of the scales if 75% of the items are non-missing.

```
items$Valued <- sjstats::mean_n(items[,Valued_vars], .75)
items$TradPed <- sjstats::mean_n(items[,TradPed_vars], .75)
items$SCRPed <- sjstats::mean_n(items[,SCRPed_vars], .75)
items$Total <- sjstats::mean_n(items, .75)
```

For each subscale, we can produce an apaTables::apa.cor.table for the items-and-OtherScales. Since there are 3 subscales, this is going to get spicy.

##### 7.10.4.1 Valued-by-the-Student Items

First with the traditional pedagogy total scale score.

```
apaTables::apa.cor.table(items[c('ValObjectives', 'IncrUnderstanding', 'IncrInterest', 'TradPed'))
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. ValObjectives	4.52	0.61			
2. IncrUnderstanding	4.28	0.82	.44**		
			[.34, .52]		
3. IncrInterest	3.94	0.99	.48**	.68**	
			[.39, .56]	[.62, .74]	
4. TradPed	4.25	0.76	.50**	.62**	.61**
			[.41, .58]	[.55, .68]	[.54, .68]

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

After each analysis, I can update my table with the data representing the correlations with the individual items of one scale with the total scale score included in the correlation matrix. Our hope is that the corrected item-total correlations that we collected above will be stronger than the individual items' correlations with the other scales' total scores.

---

#### Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50	.50	
IncrUnderstanding	.67	.62	
IncrInterest	.70	.61	
ClearResponsibilities		.79	
EffectiveAnswers		.76	
Feedback		.66	
ClearOrganization		.74	
ClearPresentation		.78	
MultPerspectives			.68
DEIntegration			.67
EquitableEval			.59
Feedback			.58

Next, the Valued-by-the-Student Items with socially responsive pedagogy total scale score.

```
apaTables::apa.cor.table(items[c('ValObjectives', 'IncrUnderstanding', 'IncrInterest', 'SCRPed
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. ValObjectives	4.52	0.61			
2. IncrUnderstanding	4.28	0.82	.44**		
			[.34, .52]		
3. IncrInterest	3.94	0.99	.48**	.68**	
			[.39, .56]	[.62, .74]	
4. SCRPed	4.52	0.58	.41**	.44**	.54**
			[.32, .50]	[.35, .53]	[.45, .61]

Note. M and SD are used to represent mean and standard deviation, respectively.

Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

---

#### Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50	.50	.41
IncrUnderstanding	.67	.62	.44
IncrInterest	.70	.61	.54
ClearResponsibilities		.79	
EffectiveAnswers		.76	
Feedback		.66	
ClearOrganization		.74	
ClearPresentation		.78	
MultPerspectives			.68
DEIntegration			.67
EquitableEval			.59
Feedback			.58

### 7.10.5 Traditional Pedagogy Items

First, the traditional pedagogy items with the valued-by-the-student total scale score.

```
apaTables::apa.cor.table(items[c('ClearResponsibilities', 'EffectiveAnswers', 'Feedback', 'ClearOrganization', 'ClearPresentation', 'Valued')])
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. ClearResponsibilities	4.44	0.82			
2. EffectiveAnswers	4.36	0.83	.69** [.63, .74]		
3. Feedback	4.24	0.88	.63** [.56, .70]	.58** [.50, .65]	
4. ClearOrganization	4.01	1.08	.67** [.61, .73]	.60** [.52, .67]	.55** [.46, .62]
5. ClearPresentation	4.24	0.92	.69** [.62, .74]	.72** [.66, .77]	.55** [.47, .62]
6. Valued	4.25	0.68	.52** [.43, .60]	.59** [.51, .66]	.49** [.40, .57]
4		5			

.70\*\*  
[.63, .75]

.63\*\* .69\*\*  
[.55, .69] [.63, .75]

Note. M and SD are used to represent mean and standard deviation, respectively.

Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

---

Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	<b>.50</b>	.50	.41
IncrUnderstanding	<b>.67</b>	.62	.44
IncrInterest	<b>.70</b>	.61	.54
ClearResponsibilities	.52	<b>.79</b>	
EffectiveAnswers	.59	<b>.76</b>	
Feedback	.49	<b>.66</b>	
ClearOrganization	.63	<b>.74</b>	
ClearPresentation	.69	<b>.78</b>	
MultPerspectives			<b>.68</b>
DEIntegration			<b>.67</b>
EquitableEval			<b>.59</b>
Feedback			<b>.58</b>

Next, the traditional pedagogy items with the socially responsive pedagogy total scale score.

```
apaTables::apa.cor.table(items[c('ClearResponsibilities', 'EffectiveAnswers', 'Feedback', 'ClearOrganization', 'MultPerspectives', 'DEIntegration', 'EquitableEval', 'IncrInterest', 'IncrUnderstanding', 'ValObjectives')])
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3
1. ClearResponsibilities	4.44	0.82			
2. EffectiveAnswers	4.36	0.83	.69**		
			[.63, .74]		
3. Feedback	4.24	0.88	.63**	.58**	
			[.56, .70]	[.50, .65]	
4. ClearOrganization	4.01	1.08	.67**	.60**	.55**
			[.61, .73]	[.52, .67]	[.46, .62]

5. ClearPresentation	4.24	0.92	.69**	.72**	.55**
			[.62, .74]	[.66, .77]	[.47, .62]
6. SCRPed	4.52	0.58	.64**	.60**	.62**
			[.56, .70]	[.52, .67]	[.55, .69]

4                5

.70\*\*  
 [.63, .75]

.51\*\*        .62\*\*  
 [.43, .59]    [.54, .68]

Note. M and SD are used to represent mean and standard deviation, respectively.  
 Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations  
 that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

---

 Item-Total Correlations of Items with their Own and Other Subscales
 

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50	.50	.41
IncrUnderstanding	.67	.62	.44
IncrInterest	.70	.61	.54
ClearResponsibilities	.52	.79	.64
EffectiveAnswers	.59	.76	.60
Feedback	.49	.66	.62
ClearOrganization	.63	.74	.51
ClearPresentation	.69	.78	.62
MultPerspectives			.68

Variables	Valued	TradPed	SCRPed
DEIintegration			.67
EquitableEval			.59
Feedback			.58

### 7.10.5.1 Socially Responsive Pedagogy Items

First, the socially responsive pedagogy items with the valued-by-the-student total scale score.

```
apaTables::apa.cor.table(items[c('MultPerspectives', 'InclusvClassrm', 'DEIintegration', 'Equitab'))
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3	4
1. MultPerspectives	4.39	0.84				
2. InclusvClassrm	4.61	0.68	.57**			
			[.49, .64]			
3. DEIintegration	4.53	0.74	.50**	.62**		
			[.41, .59]	[.54, .69]		
4. EquitableEval	4.57	0.63	.59**	.47**	.37**	
			[.51, .66]	[.37, .55]	[.27, .47]	
5. Valued	4.25	0.68	.54**	.48**	.29**	.46**
			[.45, .61]	[.38, .56]	[.18, .40]	[.37, .54]

Note. M and SD are used to represent mean and standard deviation, respectively.  
Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations  
that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

---

Item-Total Correlations of Items with their Own and Other Subscales

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50	.50	.41
IncrUnderstanding	.67	.62	.44
IncrInterest	.70	.61	.54
ClearResponsibilities	.52	.79	.64
EffectiveAnswers	.59	.76	.60
Feedback	.49	.66	.62
ClearOrganization	.63	.74	.51
ClearPresentation	.69	.78	.62
MultPerspectives	.54		.68
DEIintegration	.48		.67
EquitableEval	.29		.59
Feedback	.46		.58

Next, the socially responsive pedagogy items with the traditional pedagogy total scale score.

```
apaTables::apa.cor.table(items[c('MultPerspectives', 'InclusvClassrm', 'DEIintegration', 'Equita
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3	4
1. MultPerspectives	4.39	0.84				
2. InclusvClassrm	4.61	0.68	.57**			
			[.49, .64]			
3. DEIintegration	4.53	0.74	.50**	.62**		
			[.41, .59]	[.54, .69]		
4. EquitableEval	4.57	0.63	.59**	.47**	.37**	
			[.51, .66]	[.37, .55]	[.27, .47]	
5. TradPed	4.25	0.76	.71**	.54**	.34**	.65**
			[.64, .76]	[.46, .62]	[.23, .44]	[.58, .71]

Note. M and SD are used to represent mean and standard deviation, respectively.  
Values in square brackets indicate the 95% confidence interval.

The confidence interval is a plausible range of population correlations  
that could have caused the sample correlation (Cumming, 2014).

\* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

---

 Item-Total Correlations of Items with their Own and Other Subscales
 

---

Variables	Valued	TradPed	SCRPed
ValObjectives	.50	.50	.41
IncrUnderstanding	.67	.62	.44
IncrInterest	.70	.61	.54
ClearResponsibilities	.52	.79	.64
EffectiveAnswers	.59	.76	.60
Feedback	.49	.66	.62
ClearOrganization	.63	.74	.51
ClearPresentation	.69	.78	.62
MultPerspectives	.54	.71	.68
InclusvClassrm	.48	.54	.67
DEIintegration	.29	.34	.59
EquitableEval	.46	.65	.58

---

#### 7.10.6 APA style results section with table

- Brief description of each step
- Brief instructions for interpreting results
- Presentation of results

Item analyses were conducted on the 12 course evaluation items that were selected to represent the Valued-by-the-Student, Traditional Pedagogy, and Socially Responsive Pedagogy subscales. To assess the within-scale convergent and discriminant validity of these three subscales, each item was correlated with its own scale (with the item removed) and with the other course evaluation scales (see Table 1). For the Valued and Traditional Pedagogy dimensions, items were more highly correlated with their own scale than with the other scale. For the SCRPed subscale, two items (multiple perspectives, equitable evaluations) had higher correlations with Traditional Pedagogy than with their hypothesized subscale (Socially Responsive Pedagogy). Coefficient alphas were .77, .90, .81, and .92 for the Valued-by-the-Student, Traditional Pedagogy, Socially Responsive, and total-scale scores, respectively. We conclude that more work is needed to create distinct and stable subscales.

#### 7.10.7 Explanation to grader

# **EXPLORATORY *FACTOR* ANALYSIS**

The next two lessons are devoted to exploratory *factor* analysis. The two approaches are principal components analysis (PCA) and principal axis factoring (PAF). In truth, only PAF is considered *factor* analysis. I will explain why in the lesson.

These approaches are loosely termed *exploratory* because the statistical process (not the researcher) produces the factor (think scale or subscale) and identifies which items belong to it. This is contrasted with *confirmatory* approaches (which use structural equation modeling) where the researcher assigns items to factors and analyzes the goodness of fit.



# Chapter 8

# Principal Components Analysis

## [Screencasted Lecture Link](#)

In this lesson on principal components analysis (PCA) I provide an introduction to the exploratory factor analysis (EFA) arena. We will review the theoretical and technical aspects of PCA, we will work through a research vignette, and then consider the relationship of PCA to item analysis and reliability coefficients.

Please note, although PCA is frequently grouped into EFA techniques, it is *exploratory* but it is not *factor analysis*. We'll discuss the difference in the lecture.

## 8.1 Navigating this Lesson

There are about two hours of lecture. If you work through the materials with me, I would be plan for an additional hour-and-a-half.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

### 8.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Distinguish between PCA and PAF on several levels:
  - which path diagram represents each best, and
  - keywords associated with each: factor loadings, linear components, describe versus explain.
- Recognize/define an identity matrix – what test would you use to diagnose it?
- Recognize/define multicollinearity and singularity – what test would you use to diagnose it?

- Describe the pattern of “loadings” (i.e., the relative weights of an item on its own scale compared to other scales) that supports the structure of the instrument.
- Compare the results from item analysis and PCA.

### 8.1.2 Planning for Practice

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The least complex is to change the random seed in the research and rework the problem demonstrated in the lesson. The results *should* map onto the ones obtained in the lecture.

The second option involves utilizing one of the simulated datasets available in this OER. Szymanski and Bissonette’s [2020] Perceptions of the LGBTQ College Campus Climate Scale: Development and Psychometric Evaluation was used as the research vignette for the validity, reliability, and item analysis lesson. Keum et al.’s Gendered Racial Microaggressions Scale for Asian American Women [Keum et al., 2018] will be used in the lessons on confirmatory factor analysis. Both of these would be suitable for the PCA and PAF homework assignments.

As a third option, you are welcome to use data to which you have access and is suitable for PCA. These could include other vignettes from this OER, other simulated data, or your own data (presuming you have permission to use it). In either case, please plan to:

- Properly format and prepare the data.
- Conduct diagnostic tests to determine the suitability of the data for PCA.
- Conducting tests to guide the decisions about number of components to extract.
- Conducting orthogonal and oblique extractions (at least two each with different numbers of components).
- Selecting one solution and preparing an APA style results section (with table and figure).
- Compare your results in light of any other psychometrics lessons where you have used this data.

### 8.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

- Revelle, William. (n.d.). Chapter 6: Constructs, components, and factor models. In *An introduction to psychometric theory with applications in R*. Retrieved from <https://personality-project.org/r/book/#chapter6>
  - pp. 145 to 150 (we’ll continue with the rest in the next lecture). Stop at “6.2 Exploratory Factor Analysis.”
  - A simultaneously theoretical review of psychometric theory while working with R and data to understand the concepts.
- Revelle, W. (2019). *How To: Use the psych package for Factor Analysis and data reduction*.
  - pp. 13 through 24 provide technical information about what we are doing

- Dekay, Nicole (2021). Quick Reference Guide: The statistics for psychometrics <https://www.humanalysts.com/quick-reference-guide-the-statistics-for-psychometrics>
- Lewis, J. A., & Neville, H. A. (2015). Construction and initial validation of the Gendered Racial Microaggressions Scale for Black Women. *Journal of Counseling Psychology*, 62(2), 289–302. <https://doi.org/10.1037/cou0000062>
  - Our research vignette for this lesson.

#### 8.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(psych)){install.packages('psych')}
# if(!require(tidyverse)){install.packages('tidyverse')}
# if(!require(MASS)){install.packages('MASS')}
# if(!require(sjstats)){install.packages('sjstats')}
# if(!require(apaTables)){install.packages('apaTables')}
# if(!require(qualtrics)){install.packages('qualtrics')}
```

## 8.2 Exploratory Principal Components Analysis

The psychometric version of *parsimony* is seen in our attempt to *describe* (components) or to *explain* (factors) in the relationships between many observed variables in terms of a more limited set of components, latent factors, or dimensions.

That is, we are trying to:

- understand the structure of a set of variables,
- construct a questionnaire to measure an underlying latent variable, and
- reduce a data set to a more manageable size (e.g., representing bundles of items as subscale scores) while retaining as much of the information as possible

### 8.2.1 Some Framing Ideas (in very lay terms)

*Exploratory* versus *confirmatory* factor analysis.

- Both exploratory and confirmatory approaches to components/factor analysis are used in scale construction. Think of “scales” as being interchangeable with “factors” and “components.”
  - That said, “factors” and “components” are not interchangeable terms.

- **Exploratory:** Even though we may have an a priori model in mind, we *explore* the structure of the items by using diagnostics (KMO, Barlett's, determinant), factor extraction, and rotation to determine the number of scales (i.e., components or factors) that exist within the raw data or correlation matrix. The algorithms (including matrix algebra) determine the relationship of each item to its respective scales (i.e., components or factors).
- **Confirmatory:** Starting with an a priori theory, we specify the structure (i.e., number and levels of factors) and which items belong to factors. We use structural equation modeling as the framework. We evaluate the quality of the model with a number of fit indices.

Within the *exploratory* category we will focus on two further distinctions (there are even more). The first is principal components analysis (PCA). The second is principal axis factoring (PAF). PAF is one of the approaches that is commonly termed “exploratory factor analysis” (EFA). In this first lesson we focus on the differences between PCA and EFA.

- **Option #1/Component model:** PCA approximates the correlation matrix in terms of the product of components where each is a weighted linear sum of the variables. In the figure below, note how the arrows in the components analysis (a *path* model) point from variables to the component. Perhaps an oversimplification, think of each of these as a predictor variable contributing to an outcome.
- **Option #2/Factor model:** EFA (and in the next lesson, PAF/principal axis factoring) approximates the correlation matrix by the product of the two factors; this approach presumes that the factors are the causes (rather than as consequences). In the figure below, note how the arrows in the factor analysis model (a *structural* model) point from latent variable (or factor) to the observed variables (items). Factor analysis has been termed *causal modeling* because the latent variables are theorized to cause the responses to the individual items. There are other popular approaches, including parallel analysis (which is what the authors used in this lesson’s research vignette).

Well-crafted figures provide important clues to the analyses. In structural models, rectangles and squares indicate the presence of *observed* (also called *manifest*) variables. These are variables that have a column in the dataset. In our particular case, they are the responses to the 25 items in the GRMS.

Circles or ovals represent latent variables or factors. These were never raw data, but are composed of the relations of variables that were collected. They are more complex than mean or sum scores. Rather, they represent what the variables that represent them share in common.

Our focus today is on the principal component analysis (PCA) approach to scale construction.

### 8.3 PCA Workflow

Below is a screenshot of the workflow. The original document is located in the [Github site](#) that hosts the ReCentering Psych Stats: Psychometrics OER.

Steps in the process include:

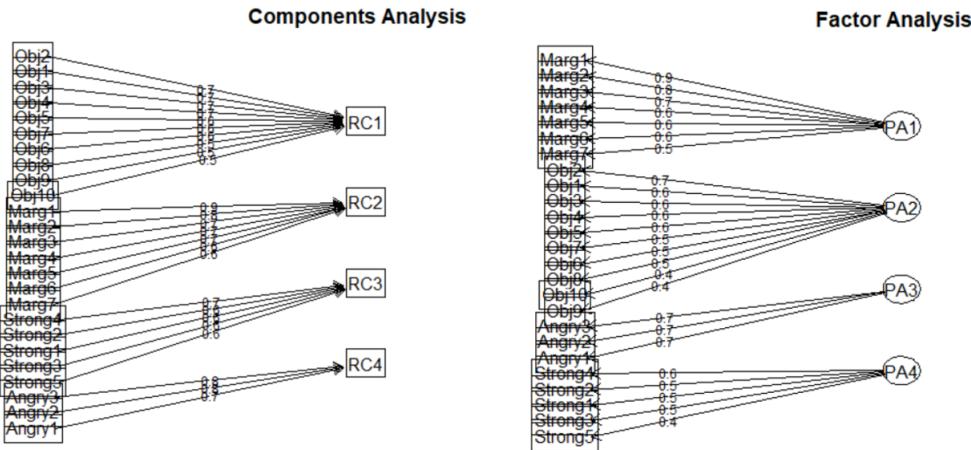


Figure 8.1: Comparison of path models for PCA and EFA for our research vignette

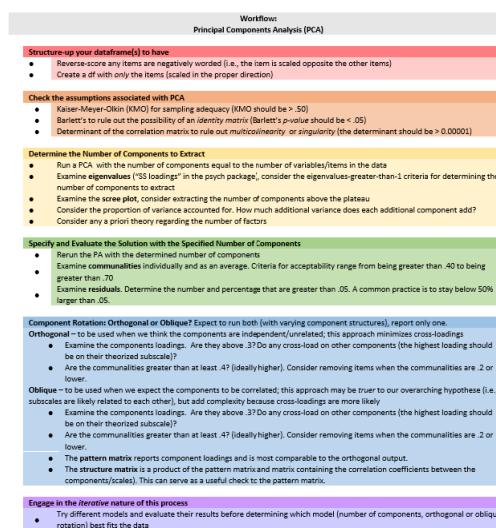


Figure 8.2: Image of the workflow for PCA

- Creating an *items only* dataframe where any items are scaled in the same direction (e.g., negatively worded items are reverse-scored).
- Conducting tests that assess the statistical assumptions of PCA to ensure that the data is appropriate for PCA.
- Determining the number of components (think “subscale”) to extract.
- Conducting the component extraction – this process will likely occur iteratively,
  - exploring orthogonal (uncorrelated/independent) and oblique (correlated) components, and
  - changing the number of components to extract

Because the intended audience for the ReCentering Psych Stats OER is the scientist-practitioner-advocate, this lesson focuses on the workflow and decisions. As you might guess, the details of PCA can be quite complex. Some important notions to consider that may not be obvious from lesson, are these:

- The values of component loadings are directly related to the correlation (similarly, the covariance) matrix between the items.
  - Although I do not explain this in detail, nearly every analytic step attempts to convey this notion by presenting equivalent analytic options using the raw data and correlation matrix.
- PCA is about *dimension reduction* – our goal is fewer components (i.e., subscales) than there are items.
  - In this lesson’s vignette there are 25 items on the scale and we will end up with 4 subscales.
- Principal component analysis is *exploratory*, but it is not “factor analysis.”
- Matrix algebra (e.g., using the transpose of a matrix, multiplying matrices together) plays a critical role in the analytic solution.

## 8.4 Research Vignette

This lesson’s research vignette emerges from Lewis and Neville’s Gendered Racial Microaggressions Scale for Black Women [2015]. The article reports on two separate studies that comprised the development, refinement, and psychometric evaluation of two parallel versions (stress appraisal, frequency) of the scale. Below, I simulate data from the final construction of the stress appraisal version as the basis of the lecture. Items were on a 6-point Likert scale ranging from 0 (*not at all stressful*) to 5 (*extremely stressful*).

Lewis and Neville [2015] reported support for a total scale score (25 items) and four subscales. Below, I list the four subscales, their number of items, and a single example item. At the outset, let me provide a content advisory. For those who hold this particular identity (or related identities) the content in the items may be upsetting. In other lessons, I often provide a variable name that gives an indication of the primary content of the item. In the case of the GRMS, I will simply provide an abbreviation of the subscale name and its respective item number. This will allow us to easily inspect the alignment of the item with its intended factor, and hopefully minimize

discomfort. If you are not a member of this particular identity, I encourage you to learn about these microaggressions by reading the article in its entirety. Please do not ask members of this group to explain why these microaggressions are harmful or ask if they have encountered them. The four factors, number of items, and sample item are as follows:

- Assumptions of Beauty and Sexual Objectification
  - 10 items
  - “Objectified me based on physical features.”
  - Abbreviated in the simulated data as “Obj#”
- Silenced and Marginalized
  - 7 items
  - “Someone has tried to ‘put me in my place.’”
  - Abbreviated in the simulated data as “Marg#”
- Strong Black Woman Stereotype
  - 5 items
  - “I have been told that I am too assertive.”
  - Abbreviated in the simulated data as “Str#”
- Angry Black Woman Stereotype
  - 3 items
  - “Someone accused me of being angry when speaking calm.”
  - Abbreviated in the simulated data as “Ang#”

Three additional scales were reported in the Lewis and Neville article [2015]. Because (a) the focus of this lesson is on exploratory factor analytic approaches and, therefore, only requires item-level data for the scale, and (b) the article does not include correlations between the subscales/scales of all involved measures, I only simulated item-level data for the GRMS items.

Below, I walk through the data simulation. This is not an essential portion of the lesson, but I will lecture it in case you are interested. None of the items are negatively worded (relative to the other items), so there is no need to reverse-score any items.

Simulating the data involved using factor loadings, means, standard deviations, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

```
# Entering the intercorrelations, means, and standard deviations from
# the journal article
```

```
LewisGRMS_generating_model <- "
  #measurement model
  Objectification =~ .69*Obj1 + .69*Obj2 + .60*Obj3 + .59*Obj4 + .55*Obj5 + .55*Obj6 + .
  Marginalized =~ .93*Marg1 + .81*Marg2 + .69*Marg3 + .67*Marg4 + .61*Marg5 + .58*Marg6 +
  Strong =~ .59*Str1 + .55*Str2 + .54*Str3 + .54*Str4 + .51*Str5
  Angry =~ .70*Ang1 + .69*Ang2 + .68*Ang3"
```

```

#Means
Objectification ~ 1.85*1
Marginalized ~ 2.67*1
Strong ~ 1.61*1
Angry ~ 2.29*1

#Correlations
Objectification ~~ .63*Marginalized
Objectification ~~ .66*Strong
Objectification ~~ .51*Angry

Marginalized ~~ .59*Strong
Marginalized ~~ .62*Angry

Strong ~~ .61*Angry

""

set.seed(240311)
dfGRMS <- lavaan::simulateData(model = LewisGRMS_generating_model, model.type = "sem",
  meanstructure = T, sample.nobs = 259, standardized = FALSE)

# used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(dfGRMS))

# The code below loops through each column of the dataframe and
# assigns the scaling accordingly Rows 1 thru 26 are the GRMS items

for (i in 1:ncol(dfGRMS)) {
  if (i >= 1 & i <= 26) {
    dfGRMS[, i] <- scales::rescale(dfGRMS[, i], c(1, 5))
  }
}

# rounding to integers so that the data resembles that which was
# collected
library(tidyverse)
dfGRMS <- dfGRMS %>%
  round(0)

# quick check of my work psych::describe(dfGRMS)

```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do). If you save the .csv file and bring it back in, you will lose any formatting (e.g., ordered factors will be interpreted as character variables).

```
#write the simulated data as a .csv
#write.table(dfGRMS, file="dfGRMS.csv", sep=",", col.names=TRUE, row.names=FALSE)
#bring back the simulated dat from a .csv file
#dfGRMS <- read.csv ("dfGRMS.csv", header = TRUE)
```

An .rds file preserves all formatting to variables prior to the export and re-import. For the purpose of this chapter, you don't need to do either. That is, you can re-simulate the data each time you work the problem.

```
# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(dfGRMS, 'dfGRMS.rds') bring back the simulated dat
# from an .rds file dfGRMS <- readRDS('dfGRMS.rds')
```

## 8.5 Working the Vignette

Below we will create a correlation matrix of our items. Whether we are conducting PCA or PAF, the *dimension-reduction* we are seeking is looking for clusters of correlated items in the *R*-matrix. Essentially, these are [Field, 2012]:

- statistical entities that can be plotted as classification axes where coordinates of variables along each axis represent the strength of the relationship between that variable to each component/factor.
- mathematical equations, resembling regression equations, where each variable is represented according to its relative weight

PCA in particular establishes which linear components exist within the data and how a particular variable might contribute to that component.

Here is the correlation matrix of our items. It would be quite a daunting exercise to visually inspect this and manually cluster the correlations of items.

```
items <- dfGRMS %>%
  dplyr::select(Obj1:Ang3)
```

Next, we create an object that holds the correlation matrix of the items.

```
GRMSmatrix <- cor(items) #correlation matrix created and saved as object
round(GRMSmatrix, 2)
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Marg1	Marg2	Marg3
Obj1	1.00	0.30	0.24	0.20	0.27	0.18	0.25	0.32	0.12	0.26	0.17	0.21	0.19
Obj2	0.30	1.00	0.32	0.24	0.27	0.21	0.24	0.29	0.26	0.19	0.08	0.19	0.14
Obj3	0.24	0.32	1.00	0.21	0.22	0.19	0.25	0.21	0.17	0.23	0.25	0.19	0.15
Obj4	0.20	0.24	0.21	1.00	0.36	0.19	0.27	0.27	0.23	0.26	0.16	0.13	0.17

Obj5	0.27	0.27	0.22	0.36	1.00	0.16	0.16	0.25	0.14	0.19	0.26	0.23	0.22
Obj6	0.18	0.21	0.19	0.19	0.16	1.00	0.16	0.19	0.14	0.10	0.16	0.06	0.05
Obj7	0.25	0.24	0.25	0.27	0.16	0.16	1.00	0.33	0.21	0.25	0.31	0.18	0.20
Obj8	0.32	0.29	0.21	0.27	0.25	0.19	0.33	1.00	0.16	0.26	0.12	0.10	0.12
Obj9	0.12	0.26	0.17	0.23	0.14	0.14	0.21	0.16	1.00	0.14	0.03	0.08	0.18
Obj10	0.26	0.19	0.23	0.26	0.19	0.10	0.25	0.26	0.14	1.00	0.10	0.10	0.20
Marg1	0.17	0.08	0.25	0.16	0.26	0.16	0.31	0.12	0.03	0.10	1.00	0.33	0.36
Marg2	0.21	0.19	0.19	0.13	0.23	0.06	0.18	0.10	0.08	0.10	0.33	1.00	0.35
Marg3	0.19	0.14	0.15	0.17	0.22	0.05	0.20	0.12	0.18	0.20	0.36	0.35	1.00
Marg4	0.21	0.15	0.20	0.24	0.21	0.13	0.21	0.17	0.07	0.17	0.41	0.20	0.37
Marg5	0.09	0.17	0.13	0.20	0.25	0.12	0.18	0.18	0.20	0.06	0.35	0.31	0.24
Marg6	0.22	0.21	0.11	0.22	0.24	0.22	0.31	0.20	0.12	0.14	0.34	0.28	0.31
Marg7	0.08	0.18	0.11	0.19	0.18	0.12	0.13	0.13	0.09	0.07	0.28	0.29	0.23
Str1	0.19	0.19	0.19	0.13	0.23	0.06	0.26	0.14	0.13	0.21	0.17	0.18	0.15
Str2	0.23	0.15	0.18	0.14	0.11	0.14	0.18	0.10	0.07	0.16	0.11	0.15	0.21
Str3	0.18	0.06	0.15	0.10	0.13	0.06	0.15	0.05	0.05	0.17	0.14	0.18	0.15
Str4	0.03	0.14	0.17	0.13	0.07	0.08	0.12	0.03	0.00	0.06	0.10	0.07	0.06
Str5	0.13	0.11	0.17	0.01	0.09	0.05	0.15	0.06	0.02	0.03	0.07	0.15	0.05
Ang1	0.06	0.01	0.15	0.14	0.11	0.04	0.25	0.08	0.12	0.06	0.21	0.19	0.13
Ang2	0.05	0.05	0.09	0.07	0.09	0.14	0.09	0.03	-0.01	0.13	0.13	0.21	0.14
Ang3	0.21	0.10	0.18	0.19	0.11	0.11	0.23	0.08	0.08	0.14	0.25	0.20	0.14
	Marg4	Marg5	Marg6	Marg7	Str1	Str2	Str3	Str4	Str5	Ang1	Ang2	Ang3	
Obj1	0.21	0.09	0.22	0.08	0.19	0.23	0.18	0.03	0.13	0.06	0.05	0.21	
Obj2	0.15	0.17	0.21	0.18	0.19	0.15	0.06	0.14	0.11	0.01	0.05	0.10	
Obj3	0.20	0.13	0.11	0.11	0.19	0.18	0.15	0.17	0.17	0.15	0.09	0.18	
Obj4	0.24	0.20	0.22	0.19	0.13	0.14	0.10	0.13	0.01	0.14	0.07	0.19	
Obj5	0.21	0.25	0.24	0.18	0.23	0.11	0.13	0.07	0.09	0.11	0.09	0.11	
Obj6	0.13	0.12	0.22	0.12	0.06	0.14	0.06	0.08	0.05	0.04	0.14	0.11	
Obj7	0.21	0.18	0.31	0.13	0.26	0.18	0.15	0.12	0.15	0.25	0.09	0.23	
Obj8	0.17	0.18	0.20	0.13	0.14	0.10	0.05	0.03	0.06	0.08	0.03	0.08	
Obj9	0.07	0.20	0.12	0.09	0.13	0.07	0.05	0.00	0.02	0.12	-0.01	0.08	
Obj10	0.17	0.06	0.14	0.07	0.21	0.16	0.17	0.06	0.03	0.06	0.13	0.14	
Marg1	0.41	0.35	0.34	0.28	0.17	0.11	0.14	0.10	0.07	0.21	0.13	0.25	
Marg2	0.20	0.31	0.28	0.29	0.18	0.15	0.18	0.07	0.15	0.19	0.21	0.20	
Marg3	0.37	0.24	0.31	0.23	0.15	0.21	0.15	0.06	0.05	0.13	0.14	0.14	
Marg4	1.00	0.27	0.28	0.24	0.13	0.17	0.13	0.16	-0.01	0.11	0.17	0.20	
Marg5	0.27	1.00	0.27	0.23	0.13	0.06	0.20	0.11	0.08	0.04	0.10	0.22	
Marg6	0.28	0.27	1.00	0.26	0.12	0.28	0.17	0.14	0.09	0.13	0.21	0.16	
Marg7	0.24	0.23	0.26	1.00	0.12	-0.01	0.05	0.05	0.03	0.18	0.12	0.08	
Str1	0.13	0.13	0.12	0.12	1.00	0.16	0.22	0.14	0.18	0.18	0.05	0.06	
Str2	0.17	0.06	0.28	-0.01	0.16	1.00	0.19	0.17	0.18	0.11	0.16	0.12	
Str3	0.13	0.20	0.17	0.05	0.22	0.19	1.00	0.27	0.19	0.27	0.13	0.22	
Str4	0.16	0.11	0.14	0.05	0.14	0.17	0.27	1.00	0.11	0.12	0.04	0.04	
Str5	-0.01	0.08	0.09	0.03	0.18	0.18	0.19	0.11	1.00	0.15	0.11	0.12	
Ang1	0.11	0.04	0.13	0.18	0.18	0.11	0.27	0.12	0.15	1.00	0.23	0.26	
Ang2	0.17	0.10	0.21	0.12	0.05	0.16	0.13	0.04	0.11	0.23	1.00	0.27	
Ang3	0.20	0.22	0.16	0.08	0.06	0.12	0.22	0.04	0.12	0.26	0.27	1.00	

This correlation matrix is so big that you might wish to write code so that you can examine it in sections

```
# round(GRMSmatrix[, 1:8], 2) round(GRMSmatrix[, 9:16], 2)
# round(GRMSmatrix[, 17:25], 2)
```

With component and factor analytic procedures we can analyze the data with either raw data or correlation matrix. Producing the matrix helps us see how this is a “structural” analysis. That is, we are trying to see if our more parsimonious extraction (i.e., our “dimension reduction”) “reproduces” this original correlation matrix. In each of the analyses I will include code for running the analyses with raw data and the  $r$ -matrix.

### 8.5.1 Three Diagnostic Tests to Evaluate the Appropriateness of the Data for Component-or-Factor Analysis

Below is a snip from the workflow to remind us where we are in the steps to PCA.



Figure 8.3: Image of an excerpt from the workflow

#### 8.5.1.1 Is my sample adequate for PCA?

There have been a number of generic guidelines (some supported by analyses, some not) about “how big” the sample size should be:

- 10-15 participants per variable
- 10 times as many participants as variables (Nunnally, 1978)
- 5 and 10 participants per variable up to 300 (Kass & Tinsley, 1979)
- 300 (Tabachnick & Fidell, 2007)
- 1000 = excellent, 300 = good, 100 = poor (Comrey & Lee, 1992)

Of course it is more complicated. Monte Carlo studies have shown that:

- if factor loadings are large (~.6), the solution is reliable regardless of size
- if communalities are large (~.6), relatively small samples (~100) are sufficient, but when they are lower (well below .5), then larger samples (>500 are indicated).

The **Kaiser-Meyer-Olkin** index (KMO) is an index of *sampling adequacy* that can be used with the actual sample to let us know if the sample size is sufficient relative to the statistical characteristics of the data. If it is below the threshold, we should probably collect more data to see if it can achieve a satisfactory value.

Kaiser’s 1974 recommendations were:

- bare minimum of .5
- values between .5 and .7 as mediocre
- values between .7 and .8 as good
- values above .9 are superb

Revelle has included a KMO test in the psych package. The function can use either raw or matrix data. Either way, the only variables in the matrix should be the items of interest. This means that everything else (e.g., total or subscale scores, ID numbers) should be removed.

```
psych::KMO(items)
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: psych::KMO(r = items)
Overall MSA =  0.84
MSA for each item =
  Obj1  Obj2  Obj3  Obj4  Obj5  Obj6  Obj7  Obj8  Obj9  Obj10 Marg1 Marg2 Marg3
  0.85  0.85  0.88  0.86  0.87  0.86  0.87  0.85  0.76  0.85  0.83  0.87  0.87
Marg4 Marg5 Marg6 Marg7 Str1 Str2 Str3 Str4 Str5 Ang1 Ang2 Ang3
  0.87  0.82  0.88  0.84  0.87  0.84  0.79  0.74  0.81  0.74  0.75  0.82
```

```
# psych::KMO(GRMSmatrix)
```

We examine the KMO values for both the overall matrix and the individual items.

At the matrix level, our  $KMO = .84$ , which falls into Kaiser's definition of *good*. You can locate this value as the "Overall MSA."

At the item level, the KMO should be  $> .50$ . Variables with values below .50 should be evaluated for exclusion from the analysis (or run the analysis with and without the variable and compare the difference). Because removing and adding variables impacts the KMO, be sure to re-evaluate the sampling adequacy if changes are made to the items (and/or sample size).

At the item level, our KMO values range between .74 and .87.

Considering both item and matrix levels, we conclude that the sample size and the data are adequate for component (or factor) analysis.

### 8.5.1.2 Are there correlations among the variables that are large enough to be analyzed?

**Bartlett's test** lets us know if a matrix is an *identity matrix*. In an identity matrix all correlation coefficients (everything on the off-diagonal) would be 0.0 (and everything on the diagonal would be 1.0).

A significant Bartlett's (i.e.,  $p < .05$ ) tells that the  $R$ -matrix is not an identity matrix. That is, there are some relationships between variables that can be analyzed.

The `cortest.bartlett()` function is in the `psych` package and can be run either from the raw data or R matrix formats.

```
psych::cortest.bartlett(items) #from the raw data
```

R was not square, finding R from data

```
$chisq  
[1] 1113.299
```

```
$df  
[1] 300
```

```
# raw data produces the warning 'R was not square, finding R from  
# data.' This means nothing other than we fed it raw data and the  
# function is creating a matrix from which to do the analysis.
```

```
# psych::cortest.bartlett(GRMSmatrix, n = 259) #if using the matrix,  
# must specify sample
```

Our Bartlett's test is significant:  $\chi^2(300) = 1113.299, p < .001$ . This means that our sample correlation matrix is statistically significantly different than an identity matrix and, therefore, supports a component-or-factor analytic approach for investigating the data.

### 8.5.1.3 Is there multicollinearity or singularity in my data?

The **determinant of the correlation matrix** should be greater than 0.00001 (that would be 4 zeros, then the 1). If it is smaller than 0.00001 then we may have an issue with *multicollinearity* (i.e., variables that are too highly correlated) or *singularity* (variables that are perfectly correlated).

The determinant function we use comes from base R. It is easiest to compute when the correlation matrix is the object. However, it is also possible to specify the command to work with the raw data.

`det(GRMSmatrix)`

```
[1] 0.01140074
```

```
# det(cor(dfGRMS))#if using the raw data
```

With a value of 0.01140, our determinant is greater than the 0.00001 requirement. If it were not, then we could identify problematic variables (i.e., those correlating too highly with others; those not correlating sufficiently with others) and re-run the diagnostic statistics.

**Summary:** Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was .84, indicating acceptable sampling adequacy. The Barlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(300) = 1113.299, p < .001$ , indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.01140, supporting the suitability of our data for analysis.

### 8.5.2 Principal Components Analysis

Below is a snip from the workflow to remind us where we are in the steps to PCA.

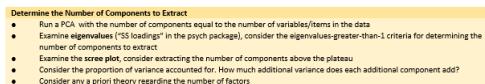


Figure 8.4: Image of an excerpt from the workflow

We can use the *principal()* function from the *psych* package with raw or matrix data.

We start by creating a principal components model that has the same number of components as there are variables in the data. This allows us to inspect the component's eigenvalues and make decisions about which to extract.

- Note, this is different than actual *factor* analysis where you *must* extract fewer factors than variables (e.g., extracting 18 [an arbitrary number] instead of 25).

```
# All of the code sets below are functionally identical. They simply
# swap out using the df or r-matrix, and whether I specify the number
# of factors or write code to instruct R to calculate it.

# pca1 <- psych::principal(GRMSmatrix, nfactors=25, rotate = 'none')
# #using the matrix form of the data and specifying the # factors

# pca1 <- psych::principal(GRMSmatrix,
# nfactors=length(GRMSmatrix[,1]), rotate = 'none') #using the matrix
# form of the data and letting the length function automatically
# calculate the # factors as a function of how many columns in the
# matrix

# pca1 <- psych::principal(items, nfactors=25, rotate='none') #using
```

```
# raw data and specifying # factors
```

```
pca1 <- psych::principal(items, nfactors = length(items), rotate = "none") # using raw data as input
```

### Principal Components Analysis

Call: psych::principal(r = items, nfactors = length(items), rotate = "none")

Standardized loadings (pattern matrix) based upon correlation matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Obj1	0.50	-0.31	0.09	-0.14	-0.18	-0.29	0.19	-0.20	-0.11	-0.02	-0.08	0.11
Obj2	0.48	-0.44	-0.04	0.14	-0.02	0.17	0.25	0.05	-0.05	0.09	0.23	0.26
Obj3	0.49	-0.21	0.18	0.06	0.02	0.08	0.03	-0.28	-0.02	-0.04	0.58	-0.07
Obj4	0.50	-0.28	-0.14	-0.09	0.14	0.14	-0.31	0.05	-0.05	0.26	-0.17	-0.10
Obj5	0.52	-0.17	-0.17	0.15	0.03	-0.06	0.07	-0.12	-0.05	0.42	-0.20	-0.38
Obj6	0.35	-0.21	-0.06	-0.28	-0.13	0.54	0.15	-0.04	0.10	-0.04	0.00	-0.31
Obj7	0.57	-0.15	0.10	-0.07	0.17	-0.07	-0.12	-0.04	0.22	-0.47	-0.15	0.08
Obj8	0.45	-0.45	-0.11	-0.12	0.06	-0.06	0.01	-0.18	0.16	-0.10	-0.27	0.26
Obj9	0.32	-0.32	-0.10	0.11	0.38	0.12	-0.03	0.55	-0.27	-0.24	0.17	-0.09
Obj10	0.42	-0.31	0.12	-0.22	-0.05	-0.37	-0.27	0.08	0.02	0.23	0.10	0.12
Marg1	0.56	0.37	-0.25	0.05	-0.04	-0.05	-0.07	-0.31	0.08	-0.26	0.07	-0.22
Marg2	0.52	0.33	-0.11	0.15	0.03	-0.18	0.35	0.07	-0.05	0.13	0.08	0.18
Marg3	0.52	0.24	-0.24	0.07	-0.19	-0.32	-0.06	0.29	-0.06	-0.09	0.18	-0.04
Marg4	0.53	0.20	-0.26	-0.06	-0.26	-0.03	-0.30	-0.13	0.01	-0.05	0.19	-0.06
Marg5	0.49	0.20	-0.29	0.24	0.08	0.18	0.05	-0.06	-0.44	-0.06	-0.22	0.03
Marg6	0.57	0.15	-0.15	-0.10	-0.27	0.13	0.09	0.23	0.13	-0.14	-0.31	0.09
Marg7	0.41	0.21	-0.38	0.15	0.19	0.13	0.10	0.05	0.39	0.20	0.06	0.28
Str1	0.42	-0.09	0.27	0.38	0.12	-0.26	-0.02	0.04	0.24	0.08	-0.06	-0.31
Str2	0.39	-0.01	0.35	-0.13	-0.49	-0.02	0.06	0.34	0.00	-0.12	-0.03	-0.13
Str3	0.39	0.22	0.47	0.18	-0.01	0.00	-0.20	0.00	-0.27	0.13	-0.21	0.12
Str4	0.27	0.07	0.34	0.37	-0.28	0.44	-0.38	-0.05	0.05	0.08	0.06	0.24
Str5	0.25	0.08	0.48	0.19	0.04	0.03	0.50	-0.10	0.04	-0.09	-0.07	-0.06
Ang1	0.37	0.31	0.33	-0.09	0.50	0.03	-0.16	0.09	0.29	-0.03	0.01	-0.04
Ang2	0.31	0.35	0.16	-0.52	0.03	0.11	0.16	0.18	0.08	0.34	0.09	0.01
Ang3	0.42	0.24	0.17	-0.41	0.28	0.01	-0.02	-0.23	-0.40	-0.05	0.03	0.08
	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24
Obj1	-0.06	-0.43	-0.01	0.17	0.09	-0.07	0.07	0.33	-0.02	0.00	0.13	-0.20
Obj2	-0.05	-0.03	-0.28	0.13	0.06	-0.12	-0.12	-0.36	-0.03	-0.19	-0.04	-0.14
Obj3	-0.05	0.06	0.08	-0.24	-0.19	0.07	-0.22	0.10	-0.12	0.22	-0.03	0.00
Obj4	-0.34	0.13	0.03	-0.14	0.21	0.14	0.18	0.03	-0.30	0.07	-0.06	-0.18
Obj5	-0.24	-0.01	-0.02	-0.03	-0.14	-0.22	-0.14	-0.04	0.18	-0.05	0.15	0.22
Obj6	0.37	-0.24	0.15	-0.07	0.05	0.00	0.24	-0.11	0.00	-0.04	-0.02	0.04
Obj7	0.02	0.18	-0.19	-0.14	0.07	-0.14	0.04	-0.05	-0.18	-0.09	0.32	0.15
Obj8	0.02	0.12	0.18	0.18	-0.42	0.14	0.09	-0.07	0.08	0.16	-0.12	0.08
Obj9	0.03	-0.05	0.09	0.12	-0.02	-0.08	-0.03	0.26	0.07	-0.02	-0.02	0.12
Obj10	0.33	0.21	0.18	-0.23	0.16	-0.03	-0.04	0.02	0.23	-0.21	-0.08	-0.03
Marg1	0.03	0.00	-0.01	-0.12	-0.01	-0.06	-0.06	0.04	0.11	-0.13	0.01	-0.17
Marg2	-0.02	-0.07	-0.13	-0.29	-0.15	0.06	0.36	0.09	-0.12	-0.16	-0.16	0.19

Marg3	-0.05	-0.02	0.21	0.05	-0.02	-0.19	0.23	-0.29	0.04	0.29	0.11	-0.11
Marg4	-0.04	0.06	0.07	0.45	0.07	0.05	-0.04	0.04	-0.18	-0.21	-0.16	0.20
Marg5	0.21	0.23	-0.01	0.02	-0.12	0.22	-0.05	0.04	0.07	-0.07	0.10	-0.23
Marg6	-0.03	-0.02	-0.10	-0.17	0.07	-0.23	-0.30	0.10	0.00	0.15	-0.32	-0.02
Marg7	0.04	-0.15	0.16	0.02	0.27	0.25	-0.16	0.05	0.08	0.10	0.20	0.09
Str1	0.34	0.01	-0.35	0.18	0.09	0.15	0.05	0.01	-0.02	0.19	-0.12	-0.03
Str2	-0.23	-0.01	-0.06	-0.05	-0.06	0.45	-0.08	-0.06	0.12	-0.08	0.12	0.04
Str3	0.22	-0.28	0.23	-0.04	-0.07	-0.04	-0.19	-0.15	-0.28	0.00	0.05	0.09
Str4	-0.07	0.05	-0.10	0.03	-0.02	-0.16	0.23	0.16	0.23	0.06	0.05	0.02
Str5	-0.14	0.34	0.38	0.13	0.26	-0.08	0.06	0.02	0.01	-0.04	-0.07	-0.01
Ang1	-0.21	-0.24	0.08	0.05	-0.21	0.01	0.00	-0.08	0.12	-0.18	-0.12	-0.19
Ang2	0.15	0.27	-0.13	0.19	-0.19	-0.13	-0.03	0.14	-0.12	0.06	0.17	-0.08
Ang3	-0.08	-0.05	-0.21	0.04	0.26	0.07	0.04	-0.12	0.23	0.19	-0.08	0.17
	PC25	h2			u2	com						
Obj1	-0.06	1	-0.000000000000000067			7.7						
Obj2	0.09	1	0.00000000000000289			7.8						
Obj3	-0.13	1	-0.00000000000000089			5.3						
Obj4	0.09	1	-0.00000000000000089			8.8						
Obj5	-0.06	1	-0.00000000000000222			7.1						
Obj6	-0.07	1	0.00000000000000100			6.8						
Obj7	-0.12	1	-0.00000000000000067			5.9						
Obj8	0.14	1	-0.00000000000000022			7.9						
Obj9	0.15	1	-0.00000000000000089			6.7						
Obj10	-0.04	1	0.00000000000000100			10.9						
Marg1	0.41	1	-0.0000000000000022			5.9						
Marg2	0.01	1	-0.0000000000000044			7.8						
Marg3	-0.07	1	-0.0000000000000022			8.2						
Marg4	-0.11	1	-0.0000000000000022			6.7						
Marg5	-0.23	1	0.0000000000000178			8.2						
Marg6	-0.09	1	-0.0000000000000022			6.8						
Marg7	0.02	1	0.0000000000000056			10.3						
Str1	0.00	1	0.0000000000000078			9.5						
Str2	0.05	1	-0.0000000000000111			6.5						
Str3	0.14	1	0.0000000000000000			9.5						
Str4	0.02	1	-0.0000000000000089			9.1						
Str5	0.01	1	-0.0000000000000067			6.2						
Ang1	-0.18	1	-0.0000000000000022			8.5						
Ang2	0.10	1	0.0000000000000022			8.1						
Ang3	0.02	1	0.0000000000000011			9.1						

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
SS loadings	5.05	1.69	1.53	1.17	1.13	1.07	1.06	0.96	0.94	0.93	0.89
Proportion Var	0.20	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
Cumulative Var	0.20	0.27	0.33	0.38	0.42	0.47	0.51	0.55	0.58	0.62	0.66
Proportion Explained	0.20	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
Cumulative Proportion	0.20	0.27	0.33	0.38	0.42	0.47	0.51	0.55	0.58	0.62	0.66
	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22
SS loadings	0.81	0.80	0.77	0.72	0.69	0.68	0.63	0.60	0.54	0.53	0.50

Proportion Var	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02
Cumulative Var	0.69	0.72	0.75	0.78	0.81	0.83	0.86	0.88	0.91	0.93	0.95
Proportion Explained	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02
Cumulative Proportion	0.69	0.72	0.75	0.78	0.81	0.83	0.86	0.88	0.91	0.93	0.95
	PC23	PC24	PC25								
SS loadings	0.47	0.45	0.42								
Proportion Var	0.02	0.02	0.02								
Cumulative Var	0.97	0.98	1.00								
Proportion Explained	0.02	0.02	0.02								
Cumulative Proportion	0.97	0.98	1.00								

Mean item complexity = 7.8

Test of the hypothesis that 25 components are sufficient.

The root mean square of the residuals (RMSR) is 0  
with the empirical chi square 0 with prob < NA

Fit based upon off diagonal values = 1

*The total variance for a particular variable will have two components: some of it will be shared with other variables (common variance,  $h^2$ ) and some of it will be specific to that measure (unique variance,  $u^2$ ). Random variance is also specific to one item, but not reliably so.*

We can examine this most easily by examining the matrix (second screen).

The columns PC1 thru PC25 are the (uninteresting at this point) unrotated loadings. PC stands for “principal component.” Although these don’t align with the specific items, at this point in the procedure, there are as many components as variables.

**Communalities** are represented as  $h^2$ . These are the proportions of common variance present in the variables. A variable that has no specific (or random) variance would have a communality of 1.0. If a variable shares none of its variance with any other variable its communality would be 0.0.

Because we extracted the same number components as variables, they all equal 1.0. That is we have explained all the variance in each variable. When we specify fewer components, the value of the communalities will decrease.

\*\*Uniquenesses\* are represented as  $u^2$ . These are the amount of unique variance for each variable. They are calculated as  $1 - h^2$  (or 1 minus the communality). Technically (at this point in the analysis where we have an equal number of components as items), they should all be zero, but the *psych* package is very “quantsy” and decimals are reported to the 15th and 16th decimal places! (hence the  $u^2$  for Q1 is -0.0000000000000006661338).

The final column, *com*, represents *item complexity*. This is an indication of how well an item reflects a single construct. If it is 1.0 then the item loads only on one component, if it is 2.0, it loads evenly on two components, and so forth. For now, we can ignore this. *I mostly wanted to reassure you that “com” is not “communality”;  $h^2$  is communality.*

Let’s switch to the first screen of output.

**Eigenvalues** are displayed in the row called *SS loadings* (i.e., the sum of squared loadings). They represent the variance explained by the particular linear component. PC1 explains 5.05 units of

variance (out of a possible 25, the total of components). As a proportion, this is  $5.05/25 = 0.20$  (reported in the *Proportion Var* row).

```
5.05/25
```

```
[1] 0.202
```

Note:

- *Cumulative Var* is helpful in determining how many components we would like to retain to balance parsimony (where the goal is frequently “as few as possible”) with the amount of variance we want to explain
- The eigenvalues are in descending order. If we were to use the *eigenvalue > 1.0* (aka, “Kaiser’s”) criteria to determine how many components to extract, we would select 7. Joliffe’s criteria was 0.7 (thus, we would select 13 components). Eigenvalues are only one criteria, let’s look at the scree plot.

*Scree plot:* We can gain another view of how many components to extract by creating a scree plot.

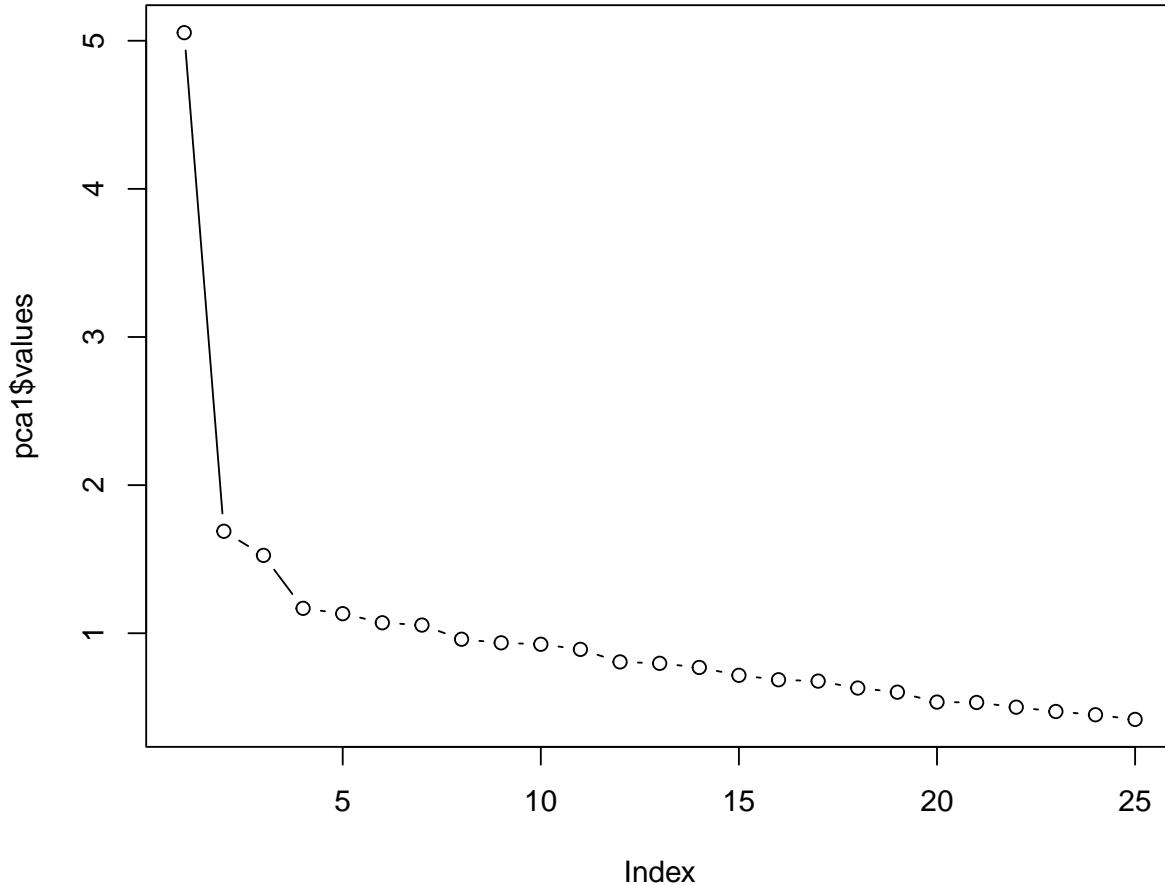
Eigenvalues are stored in the `pca1` object’s variable, “`values`”. We can see all the values captured by this object with the `names()` function:

```
names(pca1)
```

```
[1] "values"      "rotation"     "n.obs"       "communality"   "loadings"
[6] "fit"         "fit.off"      "fn"          "Call"        "uniquenesses"
[11] "complexity"  "valid"        "chi"         "EPVAL"       "R2"
[16] "objective"   "residual"     "rms"         "factors"     "dof"
[21] "null.dof"    "null.model"   "criteria"    "STATISTIC"   "PVAL"
[26] "weights"     "r.scores"     "Vaccounted" "Structure"   "scores"
```

Plotting the `eigenvalues` produces a scree plot. We can use this to further gauge the number of factors we should extract.

```
plot(pca1$values, type = "b") #type = 'b' gives us 'both' lines and points; type = 'l' gives
```



We look for the point of *inflection*. That is, where the baseline levels out into a plateau. It seems to me that there is only one clear component above the plateau. However, we see that components #5 and 5 flatten out, and then there is another drop. So it could be 1, 2, or 4.

### 8.5.3 Specifying the Number of Components

Below is a snip from the workflow to remind us where we are in the steps to PCA.

**Specify and Evaluate the Solution with the Specified Number of Components**

- Run PCA with the determined number of components
- Examine eigenvalues individually and as an average. Criteria for acceptability range from being greater than .40 to being greater than .70
- Examine residuals. Determine the number and percentage that are greater than .05. A common practice is to stay below 50% larger than .05.

Figure 8.5: Image of an excerpt from the workflow

Having determined the number of components, we re-run the analysis with this specification. Especially when researchers may not have a clear theoretical structure that guides the process, researchers may do this iteratively with varying numbers of factors. Lewis and Neville [Lewis and Neville, 2015] examined solutions with 2, 3, 4, and 5 factors. Further, they used a parallel factor analysis, whereas we used a principal components analysis).

```
# pca2 <- psych::principal(GRMSmatrix, nfactors=4, rotate='none')
pca2 <- psych::principal(items, nfactors = 4, rotate = "none") #can copy prior script, but ch
pca2
```

## Principal Components Analysis

```
Call: psych::principal(r = items, nfactors = 4, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	PC4	h2	u2	com
Obj1	0.50	-0.31	0.09	-0.14	0.37	0.63	1.9
Obj2	0.48	-0.44	-0.04	0.14	0.44	0.56	2.2
Obj3	0.49	-0.21	0.18	0.06	0.32	0.68	1.7
Obj4	0.50	-0.28	-0.14	-0.09	0.36	0.64	1.8
Obj5	0.52	-0.17	-0.17	0.15	0.35	0.65	1.6
Obj6	0.35	-0.21	-0.06	-0.28	0.25	0.75	2.7
Obj7	0.57	-0.15	0.10	-0.07	0.36	0.64	1.2
Obj8	0.45	-0.45	-0.11	-0.12	0.43	0.57	2.3
Obj9	0.32	-0.32	-0.10	0.11	0.23	0.77	2.4
Obj10	0.42	-0.31	0.12	-0.22	0.33	0.67	2.6
Marg1	0.56	0.37	-0.25	0.05	0.52	0.48	2.2
Marg2	0.52	0.33	-0.11	0.15	0.41	0.59	2.0
Marg3	0.52	0.24	-0.24	0.07	0.39	0.61	1.9
Marg4	0.53	0.20	-0.26	-0.06	0.40	0.60	1.8
Marg5	0.49	0.20	-0.29	0.24	0.42	0.58	2.6
Marg6	0.57	0.15	-0.15	-0.10	0.38	0.62	1.4
Marg7	0.41	0.21	-0.38	0.15	0.38	0.62	2.8
Str1	0.42	-0.09	0.27	0.38	0.40	0.60	2.8
Str2	0.39	-0.01	0.35	-0.13	0.29	0.71	2.2
Str3	0.39	0.22	0.47	0.18	0.46	0.54	2.7
Str4	0.27	0.07	0.34	0.37	0.33	0.67	2.9
Str5	0.25	0.08	0.48	0.19	0.34	0.66	2.0
Ang1	0.37	0.31	0.33	-0.09	0.35	0.65	3.1
Ang2	0.31	0.35	0.16	-0.52	0.51	0.49	2.7
Ang3	0.42	0.24	0.17	-0.41	0.43	0.57	2.9

	PC1	PC2	PC3	PC4
SS loadings	5.05	1.69	1.53	1.17
Proportion Var	0.20	0.07	0.06	0.05
Cumulative Var	0.20	0.27	0.33	0.38
Proportion Explained	0.54	0.18	0.16	0.12
Cumulative Proportion	0.54	0.71	0.88	1.00

Mean item complexity = 2.3

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07

```
Fit based upon off diagonal values = 0.86
```

Our eigenvalues/SS loadings remain the same. With 4 components, we explain 38% of the variance (we can see this in the “Cumulative Var” row).

*Communality* is the proportion of common variance within a variable. Principal components analysis assumes that all variance is common. Before extraction, all variance was set at 1.0, therefore, changing from 25 to 4 components will change this value ( $h^2$ ) as well as its associated *uniqueness* ( $u^2$ ), which is calculated as “1.0 minus the communality.”

The *communalities* ( $h^2$ ) and *uniquenesses* ( $u^2$ ) have changed.

Now we see that 37% of the variance associate with Obj1 is common/shared (the  $h^2$  value).

Recall that we could represent this scale with all 25 items as components, but we want a more *parsimonious* explanation. By respecifying a smaller number of components, we lose some information. That is, the retained components (now 4) cannot explain all of the variance present in the data (as we saw, it explains about 38%, cumulatively). The amount of variance explained in each variable is represented by the communalities after extraction.

We can examine the communalities through the lens of Kaiser’s criterion (the eigenvalue  $> 1$  criteria) to see if we think that “four” was a good number of components to extract.

Kaiser’s criterion is believed to be accurate if:

- when there are fewer than 30 variables (we had 25) and, after extraction, the communalities are greater than .70
  - looking at our data, none are  $> .70$ , so, this does not support extracting four components
- when the sample size is greater than 250 (ours was 259) and the average communality is  $> .60$ 
  - we can extract the communalities from our object and calculate the mean the average communality

Using the *names()* function again, we see that “communality” is available. Thus, we can easily calculate their mean. To get this value let’s first examine the possible contents of the object we created from this PCA analysis by asking for its names.

```
names(pca2)
```

```
[1] "values"      "rotation"     "n.obs"       "communality"   "loadings"
[6] "fit"         "fit.off"      "fn"          "Call"        "uniquenesses"
[11] "complexity"  "valid"        "chi"         "EPVAL"       "R2"
[16] "objective"   "residual"    "rms"         "factors"     "dof"
[21] "null.dof"    "null.model"  "criteria"    "STATISTIC"   "PVAL"
[26] "weights"     "r.scores"    "Vaccounted" "Structure"   "scores"
```

We see that it includes communalities. Thus, we can easily calculate their mean.

```
mean(pca2$communality)
```

```
[1] 0.3774861
```

```
# sum(pca2$communality) #checking my work by calculating the sum and
# dividing by 25 12.14492/25
```

We see that the average communality is 0.38. These two criteria would suggest that we may not have the best solution. That said (in our defense):

- We used the scree plot as a guide and there was support for four dimensions.
- We have an adequate sample size and that was supported with the KMO.
- Are the number of components consistent with theory? We have not yet inspected the component loadings. This will provide us with more information.

We could do several things:

- re-run with a different number of components (recall Lewis and Neville [2015] ran models with 2, 3, 4, and 5 factors)
- conduct more diagnostics
  - reproduced correlation matrix
  - the difference between the reproduced correlation matrix and the correlation matrix in the data

The *factor.model()* function in *psych* produces the *reproduced correlation matrix* by using the *loadings* from our extracted object. Conceptually, this matrix is the correlations that should be produced if we did not have the raw data but we only had the component loadings. We could do fancy matrix algebra and produce these.

The questions, though, is: How close did we get? How different is the *reproduced correlation matrix* from *GRMSmatrix* – the *R*-matrix produced from our raw data.

```
# produces the reproduced correlation matrix
round(psych::factor.model(pca2$loadings), 3)
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Marg1
Obj1	0.369	0.350	0.318	0.334	0.275	0.271	0.345	0.366	0.236	0.343	0.136
Obj2	0.350	0.442	0.329	0.356	0.352	0.222	0.320	0.398	0.317	0.300	0.125
Obj3	0.318	0.329	0.321	0.277	0.271	0.188	0.323	0.288	0.216	0.278	0.159
Obj4	0.334	0.356	0.277	0.356	0.320	0.265	0.317	0.374	0.257	0.298	0.212
Obj5	0.275	0.352	0.271	0.320	0.351	0.186	0.291	0.312	0.259	0.218	0.280
Obj6	0.271	0.222	0.188	0.265	0.186	0.249	0.242	0.289	0.155	0.264	0.120
Obj7	0.345	0.320	0.323	0.317	0.291	0.242	0.356	0.316	0.212	0.309	0.235
Obj8	0.366	0.398	0.288	0.374	0.312	0.289	0.316	0.425	0.288	0.336	0.111
Obj9	0.236	0.317	0.216	0.257	0.259	0.155	0.212	0.288	0.234	0.199	0.094



Obj6	0.070	0.173	0.204
Obj7	0.203	0.180	0.251
Obj8	0.002	0.027	0.113
Obj9	-0.025	-0.087	-0.004
Obj10	0.117	0.154	0.212
Marg1	0.232	0.238	0.262
Marg2	0.241	0.180	0.215
Marg3	0.179	0.170	0.207
Marg4	0.175	0.225	0.252
Marg5	0.121	0.049	0.103
Marg6	0.214	0.258	0.291
Marg7	0.077	0.064	0.097
Str1	0.179	-0.056	0.042
Str2	0.267	0.242	0.276
Str3	0.350	0.180	0.222
Str4	0.200	-0.024	0.039
Str5	0.257	0.084	0.128
Ang1	0.345	0.322	0.321
Ang2	0.322	0.514	0.456
Ang3	0.321	0.456	0.434

We're not really interested in this matrix. We just need it to compare it to the *GRMSmatrix* to produce the residuals. We do that next.

**Residuals** are the difference between the reproduced (i.e., those created from our component loadings) and *R*-matrix produced by the raw data.

If we look at the  $r_{Obj1Obj2}$  in our original correlation matrix (theoretically from the raw data [although we simulated data]), the value is 0.30. The reproduced correlation that we just calculated for this pair is 0.350. The difference is -0.05.

```
0.3 - 0.35
```

```
[1] -0.05
```

By using the *factor.residuals()* function R will calculate the residuals for each pair.

```
round(psych::factor.residuals(GRMSmatrix, pca2$loadings), 3)
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10
Obj1	0.631	-0.048	-0.076	-0.137	-0.006	-0.095	-0.092	-0.049	-0.120	-0.085
Obj2	-0.048	0.558	-0.013	-0.111	-0.079	-0.013	-0.079	-0.113	-0.054	-0.109
Obj3	-0.076	-0.013	0.679	-0.071	-0.051	0.000	-0.071	-0.077	-0.046	-0.048
Obj4	-0.137	-0.111	-0.071	0.644	0.038	-0.078	-0.043	-0.100	-0.032	-0.038
Obj5	-0.006	-0.079	-0.051	0.038	0.649	-0.030	-0.126	-0.063	-0.117	-0.033
Obj6	-0.095	-0.013	0.000	-0.078	-0.030	0.751	-0.084	-0.100	-0.016	-0.162
Obj7	-0.092	-0.079	-0.071	-0.043	-0.126	-0.084	0.644	0.012	-0.005	-0.058

Obj8	-0.049	-0.113	-0.077	-0.100	-0.063	-0.100	0.012	0.575	-0.131	-0.080
Obj9	-0.120	-0.054	-0.046	-0.032	-0.117	-0.016	-0.005	-0.131	0.766	-0.063
Obj10	-0.085	-0.109	-0.048	-0.038	-0.033	-0.162	-0.058	-0.080	-0.063	0.671
Marg1	0.033	-0.046	0.093	-0.056	-0.024	0.039	0.072	0.013	-0.065	0.014
Marg2	0.082	0.065	0.014	-0.043	-0.021	-0.018	-0.042	0.020	-0.012	0.033
Marg3	0.039	-0.026	-0.017	-0.055	-0.061	-0.072	-0.028	-0.028	0.051	0.099
Marg4	0.020	-0.016	0.033	-0.015	-0.065	-0.045	-0.038	-0.014	-0.057	0.026
Marg5	-0.029	-0.017	-0.030	-0.007	-0.054	0.047	-0.017	0.049	0.044	-0.001
Marg6	-0.013	0.011	-0.102	-0.051	-0.036	0.015	0.023	-0.020	-0.021	-0.055
Marg7	0.000	0.044	0.010	0.002	-0.082	0.046	-0.017	0.021	-0.026	0.045
Str1	-0.015	-0.089	-0.106	-0.030	-0.006	0.023	0.015	-0.015	-0.047	0.058
Str2	-0.021	-0.010	-0.067	-0.025	-0.014	-0.015	-0.086	-0.053	-0.012	-0.072
Str3	0.032	-0.034	-0.084	0.048	0.017	0.047	-0.072	0.046	0.026	0.058
Str4	-0.060	0.007	-0.027	0.095	-0.059	0.125	-0.030	0.022	-0.069	0.003
Str5	0.007	0.019	-0.038	-0.016	0.030	0.062	-0.020	0.056	-0.007	-0.070
Ang1	-0.070	0.000	-0.021	0.080	0.040	-0.030	0.043	0.081	0.150	-0.057
Ang2	-0.088	0.129	0.010	-0.009	0.088	-0.029	-0.091	0.006	0.074	-0.020
Ang3	0.002	0.070	0.016	0.030	0.022	-0.098	-0.021	-0.033	0.080	-0.075
	Marg1	Marg2	Marg3	Marg4	Marg5	Marg6	Marg7	Str1	Str2	Str3
Obj1	0.033	0.082	0.039	0.020	-0.029	-0.013	0.000	-0.015	-0.021	0.032
Obj2	-0.046	0.065	-0.026	-0.016	-0.017	0.011	0.044	-0.089	-0.010	-0.034
Obj3	0.093	0.014	-0.017	0.033	-0.030	-0.102	0.010	-0.106	-0.067	-0.084
Obj4	-0.056	-0.043	-0.055	-0.015	-0.007	-0.051	0.002	-0.030	-0.025	0.048
Obj5	-0.024	-0.021	-0.061	-0.065	-0.054	-0.036	-0.082	-0.006	-0.014	0.017
Obj6	0.039	-0.018	-0.072	-0.045	0.047	0.015	0.046	0.023	-0.015	0.047
Obj7	0.072	-0.042	-0.028	-0.038	-0.017	0.023	-0.017	0.015	-0.086	-0.072
Obj8	0.013	0.020	-0.028	-0.014	0.049	-0.020	0.021	-0.015	-0.053	0.046
Obj9	-0.065	-0.012	0.051	-0.057	0.044	-0.021	-0.026	-0.047	-0.012	0.026
Obj10	0.014	0.033	0.099	0.026	-0.001	-0.055	0.045	0.058	-0.072	0.058
Marg1	0.481	-0.117	-0.088	-0.023	-0.081	-0.069	-0.132	0.019	-0.012	-0.051
Marg2	-0.117	0.592	-0.036	-0.162	-0.073	-0.060	-0.051	-0.030	0.012	-0.070
Marg3	-0.088	-0.036	0.605	-0.011	-0.144	-0.051	-0.135	-0.008	0.098	-0.002
Marg4	-0.023	-0.162	-0.011	0.605	-0.090	-0.098	-0.107	0.018	0.050	0.015
Marg5	-0.081	-0.073	-0.144	-0.090	0.584	-0.052	-0.153	-0.064	0.007	0.066
Marg6	-0.069	-0.060	-0.051	-0.098	-0.052	0.622	-0.042	-0.027	0.099	0.008
Marg7	-0.132	-0.051	-0.135	-0.107	-0.153	-0.042	0.619	0.012	-0.016	-0.002
Str1	0.019	-0.030	-0.008	0.018	-0.064	-0.027	0.012	0.601	-0.051	-0.118
Str2	-0.012	0.012	0.098	0.050	0.007	0.099	-0.016	-0.051	0.706	-0.099
Str3	-0.051	-0.070	-0.002	0.015	0.066	0.008	-0.002	-0.118	-0.099	0.541
Str4	-0.013	-0.115	-0.040	0.114	-0.026	0.060	0.002	-0.194	-0.006	-0.083
Str5	0.015	0.017	0.001	-0.019	0.037	0.027	0.068	-0.127	-0.067	-0.190
Ang1	-0.018	-0.054	-0.049	-0.061	-0.078	-0.086	0.108	0.005	-0.154	-0.084
Ang2	-0.113	0.029	-0.032	-0.053	0.048	-0.044	0.053	0.108	-0.080	-0.055
Ang3	-0.013	-0.015	-0.069	-0.056	0.118	-0.129	-0.017	0.021	-0.159	-0.004
	Str4	Str5	Ang1	Ang2	Ang3					
Obj1	-0.060	0.007	-0.070	-0.088	0.002					
Obj2	0.007	0.019	0.000	0.129	0.070					
Obj3	-0.027	-0.038	-0.021	0.010	0.016					

Obj4	0.095	-0.016	0.080	-0.009	0.030
Obj5	-0.059	0.030	0.040	0.088	0.022
Obj6	0.125	0.062	-0.030	-0.029	-0.098
Obj7	-0.030	-0.020	0.043	-0.091	-0.021
Obj8	0.022	0.056	0.081	0.006	-0.033
Obj9	-0.069	-0.007	0.150	0.074	0.080
Obj10	0.003	-0.070	-0.057	-0.020	-0.075
Marg1	-0.013	0.015	-0.018	-0.113	-0.013
Marg2	-0.115	0.017	-0.054	0.029	-0.015
Marg3	-0.040	0.001	-0.049	-0.032	-0.069
Marg4	0.114	-0.019	-0.061	-0.053	-0.056
Marg5	-0.026	0.037	-0.078	0.048	0.118
Marg6	0.060	0.027	-0.086	-0.044	-0.129
Marg7	0.002	0.068	0.108	0.053	-0.017
Str1	-0.194	-0.127	0.005	0.108	0.021
Str2	-0.006	-0.067	-0.154	-0.080	-0.159
Str3	-0.083	-0.190	-0.084	-0.055	-0.004
Str4	0.672	-0.195	-0.076	0.062	0.001
Str5	-0.195	0.662	-0.109	0.027	-0.012
Ang1	-0.076	-0.109	0.655	-0.094	-0.057
Ang2	0.062	0.027	-0.094	0.486	-0.191
Ang3	0.001	-0.012	-0.057	-0.191	0.566

There are several strategies to evaluate this matrix:

- See how large the residuals are compared to the original correlations.
  - The worst possible model would occur if we extracted no components and would be the size of the original correlations.
  - If the correlations were small to start with, we expect small residuals.
  - If the correlations were large to start with, the residuals will be relatively larger (this is not terribly problematic).
- Comparing residuals requires squaring them first (because residuals can be both positive and negative).
  - The sum of the squared residuals divided by the sum of the squared correlations is an estimate of model fit. Subtracting this from 1.0 means that it ranges from 0 to 1. Values > .95 are an indication of good fit.

Analyzing the residuals means we need to extract only the upper right of the triangle of the matrix into an object. We can do this in steps.

```
# first extract the residuals
pca2_resids <- psych::factor.residuals(GRMSmatrix, pca2$loadings)
# the object has the residuals in a single column
pca2_resids <- as.matrix(pca2_resids[upper.tri(pca2_resids)])
# display the first 6 rows of the residuals
head(pca2_resids)
```

```
[,1]
[1,] -0.04849426
[2,] -0.07636484
[3,] -0.01342834
[4,] -0.13732540
[5,] -0.11134378
[6,] -0.07134501
```

One criteria of residual analysis is to see how many residuals there are that are greater than an absolute value of 0.05. The result will be a single column with TRUE if it is  $> |0.05|$  and false if it is smaller. The sum function will tell us how many TRUE responses are in the matrix. Further, we can write script to obtain the proportion of total number of residuals.

```
large.resid <- abs(pca2_resids) > 0.05
# large.resid
sum(large.resid)
```

```
[1] 139
```

```
round(sum(large.resid)/nrow(pca2_resids), 3)
```

```
[1] 0.463
```

We learn that there are 107 residuals greater than the absolute value of 0.05. This represents 36% of the total number of residuals.

There are no hard rules about what proportion of residuals can be greater than 0.05. A common practice is to stay below 50% [Field, 2012].

Another approach to analyzing residuals is to look at their mean. Because of the +/- valences, we need to square them (to eliminate the negative), take the average, then take the square root.

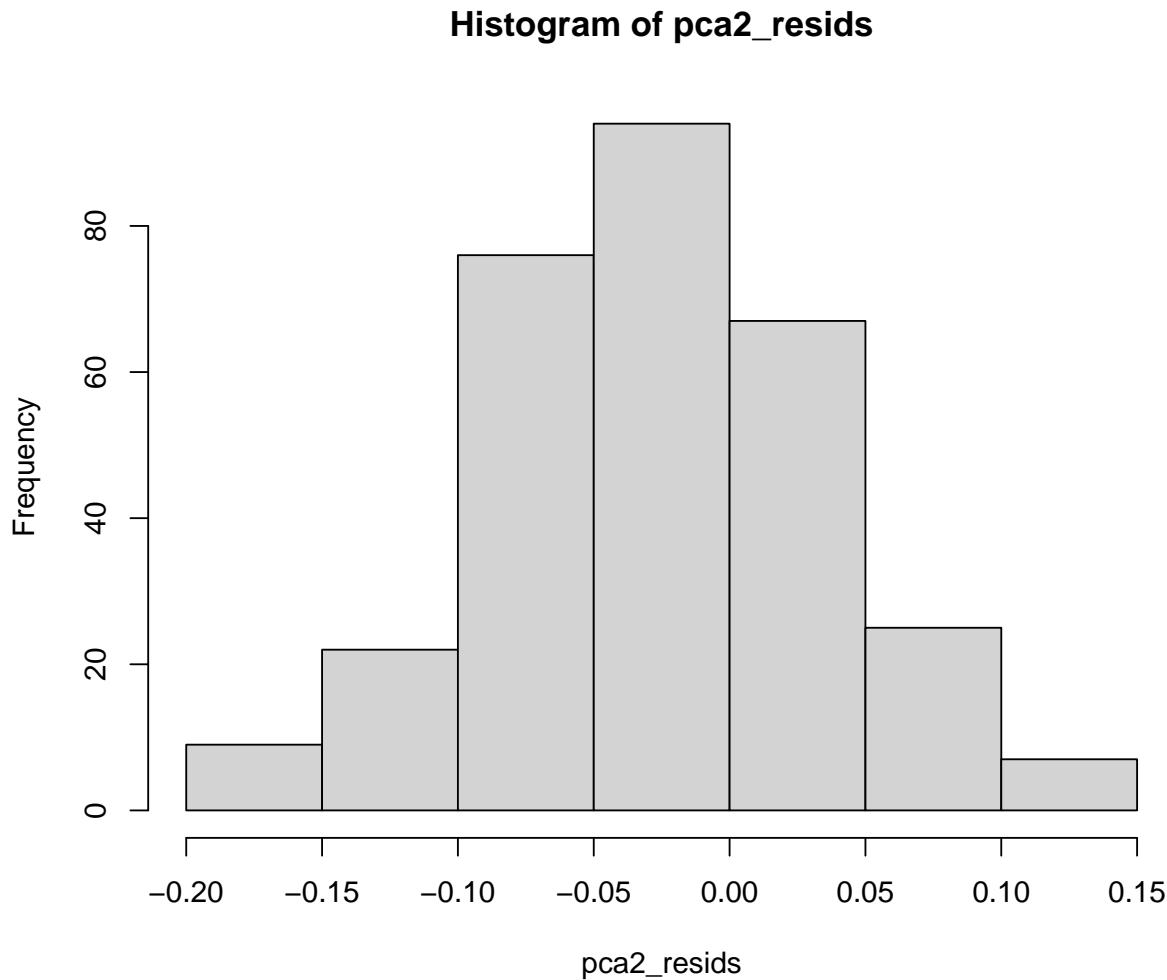
```
round(sqrt(mean(pca2_resids^2)), 3)
```

```
[1] 0.067
```

While there are no clear guidelines to interpret these, one recommendation is to consider extracting more components if the value is higher than 0.08 [Field, 2012]. Our value of 0.067 is  $< 0.08$ .

Finally, we expect our residuals to be normally distributed. A histogram can help us inspect the distribution.

```
hist(pca2_resids)
```



This looks reasonably normal to me and I do not see an indication of outliers.

#### 8.5.3.1 Quick recap of how to evaluate the # of components we extracted

- If fewer than 30 variables, the eigenvalue  $> 1$  (Kaiser's) critera is fine, so long as communalities are all  $> .70$ .
- If sample size  $> 250$  and the average communalities are  $.6$  or greater, this is fine.
- When  $N > 200$ , the scree plot can be used.
- Regarding residuals
  - fewer than 50% should have absolute values  $> 0.05$
  - model fit should be  $> 0.90$

#### 8.5.4 Component Rotation

Below is a snip from the workflow to remind us where we are in the steps to PCA.

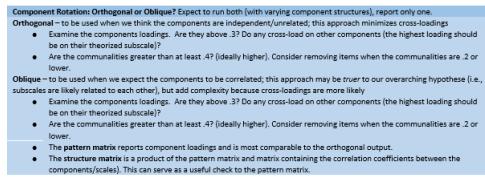


Figure 8.6: Image of an excerpt from the workflow

Rotation improves the interpretation of the components by maximizing the loading on each variable on one of the extracted components while minimizing the loading on all other components. Rotation works by changing the absolute values of the variables while keeping their differential values constant.

There are two big choices and we need to make them on theoretical grounds:

- Orthogonal rotation if you think that the components are independent/unrelated.
  - Varimax is the most common orthogonal rotation.
- Oblique rotation if you think that the components are related correlated.
  - Oblimin and promax are common oblique rotations.

Which to do?

- Orthogonal is sometimes considered to be “easier” because it minimizes cross-loadings, but
- Can you think of a measure where the subscales would *not* be correlated?

#### 8.5.4.1 Orthogonal rotation

```
# pcaORTH <- psych::principal(GRMSmatrix, nfactors = 4, rotate =
# 'varimax')
pcaORTH <- psych::principal(items, nfactors = 4, rotate = "varimax")
pcaORTH
```

```
Principal Components Analysis
Call: psych::principal(r = items, nfactors = 4, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
    RC2   RC1   RC3   RC4   h2   u2 com
Obj1  0.56  0.07  0.13  0.17  0.37  0.63  1.3
Obj2  0.61  0.14  0.16 -0.15  0.44  0.56  1.4
Obj3  0.45  0.12  0.31  0.08  0.32  0.68  2.0
Obj4  0.55  0.23  0.00  0.07  0.36  0.64  1.4
Obj5  0.44  0.37  0.12 -0.08  0.35  0.65  2.2
Obj6  0.43  0.08 -0.10  0.23  0.25  0.75  1.7
Obj7  0.47  0.20  0.22  0.21  0.36  0.64  2.3
```

Obj8	0.64	0.09	-0.04	0.01	0.43	0.57	1.1
Obj9	0.44	0.13	0.06	-0.15	0.23	0.77	1.5
Obj10	0.52	-0.01	0.08	0.22	0.33	0.67	1.4
Marg1	0.08	0.68	0.09	0.20	0.52	0.48	1.2
Marg2	0.05	0.58	0.23	0.14	0.41	0.59	1.5
Marg3	0.15	0.59	0.08	0.12	0.39	0.61	1.2
Marg4	0.20	0.56	-0.01	0.20	0.40	0.60	1.5
Marg5	0.13	0.62	0.12	-0.06	0.42	0.58	1.2
Marg6	0.27	0.48	0.06	0.27	0.38	0.62	2.2
Marg7	0.09	0.61	-0.03	-0.03	0.38	0.62	1.0
Str1	0.25	0.15	0.55	-0.12	0.40	0.60	1.7
Str2	0.26	0.00	0.34	0.34	0.29	0.71	2.9
Str3	0.02	0.12	0.63	0.23	0.46	0.54	1.3
Str4	0.04	0.10	0.56	-0.06	0.33	0.67	1.1
Str5	0.04	-0.03	0.57	0.12	0.34	0.66	1.1
Ang1	-0.01	0.17	0.37	0.42	0.35	0.65	2.3
Ang2	0.01	0.14	0.01	0.70	0.51	0.49	1.1
Ang3	0.14	0.18	0.09	0.61	0.43	0.57	1.3

	RC2	RC1	RC3	RC4
SS loadings	2.99	2.88	1.90	1.67
Proportion Var	0.12	0.12	0.08	0.07
Cumulative Var	0.12	0.23	0.31	0.38
Proportion Explained	0.32	0.30	0.20	0.18
Cumulative Proportion	0.32	0.62	0.82	1.00

Mean item complexity = 1.6

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07

Fit based upon off diagonal values = 0.86

Essentially, we have the same information as before, except that loadings are calculated after rotation (which adjusts the absolute values of the component loadings while keeping their differential values constant). Our communality and uniqueness values remain the same. The eigenvalues (SS loadings) should even out, but the proportion of variance explained and cumulative variance will remain the same (38%).

The `print.psych()` function facilitates interpretation and prioritizes the information about which we care most:

- “cut” will display loadings above .3
    - if some items load on no factors
    - if some items have cross-loadings (and their relative weights)
  - “sort” will reorder the loadings to make it clearer (to the best of its ability...in the case of ties) to which component/scale it belongs

```
pca_table <- psych::print.psych(pcaORTH, cut = 0.3, sort = TRUE)
```

## Principal Components Analysis

```
Call: psych::principal(r = items, nfactors = 4, rotate = "varimax")
```

Standardized loadings (pattern matrix) based upon correlation matrix

item	RC2	RC1	RC3	RC4	h2	u2	com
Obj8	8	0.64			0.43	0.57	1.1
Obj2	2	0.61			0.44	0.56	1.4
Obj1	1	0.56			0.37	0.63	1.3
Obj4	4	0.55			0.36	0.64	1.4
Obj10	10	0.52			0.33	0.67	1.4
Obj7	7	0.47			0.36	0.64	2.3
Obj3	3	0.45	0.31		0.32	0.68	2.0
Obj5	5	0.44	0.37		0.35	0.65	2.2
Obj9	9	0.44			0.23	0.77	1.5
Obj6	6	0.43			0.25	0.75	1.7
Marg1	11	0.68			0.52	0.48	1.2
Marg5	15	0.62			0.42	0.58	1.2
Marg7	17	0.61			0.38	0.62	1.0
Marg3	13	0.59			0.39	0.61	1.2
Marg2	12	0.58			0.41	0.59	1.5
Marg4	14	0.56			0.40	0.60	1.5
Marg6	16	0.48			0.38	0.62	2.2
Str3	20		0.63		0.46	0.54	1.3
Str5	22		0.57		0.34	0.66	1.1
Str4	21		0.56		0.33	0.67	1.1
Str1	18		0.55		0.40	0.60	1.7
Ang2	24			0.70	0.51	0.49	1.1
Ang3	25			0.61	0.43	0.57	1.3
Ang1	23		0.37	0.42	0.35	0.65	2.3
Str2	19		0.34	0.34	0.29	0.71	2.9

	RC2	RC1	RC3	RC4
SS loadings	2.99	2.88	1.90	1.67
Proportion Var	0.12	0.12	0.08	0.07
Cumulative Var	0.12	0.23	0.31	0.38
Proportion Explained	0.32	0.30	0.20	0.18
Cumulative Proportion	0.32	0.62	0.82	1.00

Mean item complexity = 1.6

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07

Fit based upon off diagonal values = 0.86

In the unrotated solution, most variables loaded on the first component. After rotation, there are four clear components/scales. Further, there is clear (or at least reasonable) component/scale membership for each item. This table lists all factor loadings that are greater than 0.30. When an item has multiple factor loadings listed, we inspect it for “cross-loading.” We observe cross-loadings with the following items: Obj3, Obj5, Ang1, Str2.

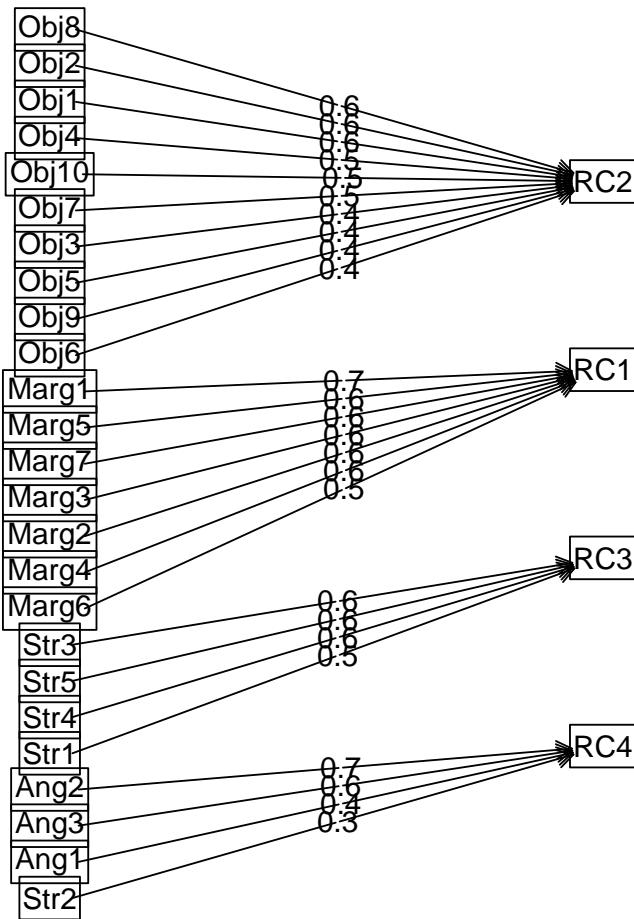
If this were a new scale and we had not yet established ideas for subscales, the next step would be to examine the items, themselves, and try to name the scales/components. If our scale construction included a priori/planned subscales, this is where we hope that the items fall where they were hypothesized to do so. Our simulated data worked pretty well, and with the exception of one item (i.e., Str2) replicated the four scales that Lewis and Neville [2015] reported in the article.

- Assumptions of Beauty and Sexual Objectification
- Silenced and Marginalized
- Strong Woman Stereotype
- Angry Woman Stereotype

We can also create a figure of the result.

```
psych::fa.diagram(pcaORTH)
```

## Components Analysis



We can extract the component loadings and write them to a table. This can be useful in preparing an APA style table for a manuscript or presentation.

```
names(pcaORTH)

[1] "values"      "rotation"     "n.obs"       "communality"   "loadings"
[6] "fit"         "fit.off"      "fn"          "Call"         "uniquenesses"
[11] "complexity"  "valid"        "chi"         "EPVAL"        "R2"
[16] "objective"   "residual"     "rms"         "factors"      "dof"
[21] "null.dof"    "null.model"   "criteria"    "STATISTIC"    "PVAL"
[26] "weights"     "r.scores"     "rot.mat"     "Vaccounted"   "Structure"
[31] "scores"

pcaORTH_table <- round(pcaORTH$loadings, 3)
write.table(pcaORTH_table, file = "pcaORTH_table.csv", sep = ",",
            row.names = FALSE)
pcaORTH_table
```

Loadings:

	RC2	RC1	RC3	RC4
Obj1	0.563		0.135	0.172
Obj2	0.613	0.137	0.158	-0.151
Obj3	0.450	0.116	0.314	
Obj4	0.545	0.232		
Obj5	0.441	0.370	0.116	
Obj6	0.426			0.228
Obj7	0.471	0.200	0.224	0.209
Obj8	0.644			
Obj9	0.437	0.127		-0.154
Obj10	0.524			0.216
Marg1		0.682		0.198
Marg2		0.576	0.234	0.136
Marg3	0.146	0.594		0.119
Marg4	0.202	0.559		0.204
Marg5	0.130	0.619	0.116	
Marg6	0.266	0.483		0.265
Marg7		0.610		
Str1	0.252	0.154	0.547	-0.117
Str2	0.258		0.335	0.339
Str3		0.124	0.627	0.226
Str4		0.101	0.559	
Str5			0.566	0.121
Ang1		0.167	0.373	0.422
Ang2		0.144		0.702
Ang3	0.142	0.177		0.611

	RC2	RC1	RC3	RC4
SS loadings	2.99	2.879	1.900	1.667
Proportion Var	0.12	0.115	0.076	0.067
Cumulative Var	0.12	0.235	0.311	0.377

#### 8.5.4.2 Oblique rotation

Whereas the orthogonal rotation sought to maximize the independence/unrelatedness of the components, an oblique rotation will allow them to be correlated. Researchers often explore both solutions but then only report one.

```
# pcaOBL <- psych::principal(GRMSmatrix, nfactors = 4, rotate =
# 'oblimin')
pcaOBL <- psych::principal(items, nfactors = 4, rotate = "oblimin")
```

Loading required namespace: GPArotation

```
pcaOBL
```

### Principal Components Analysis

Call: psych::principal(r = items, nfactors = 4, rotate = "oblimin")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	TC2	TC1	TC3	TC4	h2	u2	com
Obj1	0.57	-0.04	0.08	0.13	0.37	0.63	1.2
Obj2	0.61	0.05	0.10	-0.21	0.44	0.56	1.3
Obj3	0.43	0.02	0.27	0.02	0.32	0.68	1.7
Obj4	0.54	0.16	-0.08	0.02	0.36	0.64	1.2
Obj5	0.39	0.33	0.05	-0.15	0.35	0.65	2.3
Obj6	0.45	0.00	-0.15	0.22	0.25	0.75	1.7
Obj7	0.44	0.10	0.17	0.15	0.36	0.64	1.6
Obj8	0.67	0.00	-0.11	-0.02	0.43	0.57	1.1
Obj9	0.43	0.08	0.01	-0.19	0.23	0.77	1.5
Obj10	0.55	-0.12	0.04	0.19	0.33	0.67	1.3
Marg1	-0.05	0.69	0.02	0.13	0.52	0.48	1.1
Marg2	-0.07	0.58	0.18	0.06	0.41	0.59	1.2
Marg3	0.04	0.60	0.01	0.05	0.39	0.61	1.0
Marg4	0.12	0.55	-0.08	0.15	0.40	0.60	1.3
Marg5	0.02	0.64	0.05	-0.13	0.42	0.58	1.1
Marg6	0.19	0.45	-0.01	0.21	0.38	0.62	1.8
Marg7	-0.01	0.65	-0.09	-0.09	0.38	0.62	1.1
Str1	0.19	0.09	0.53	-0.21	0.40	0.60	1.6
Str2	0.24	-0.10	0.32	0.29	0.29	0.71	3.1
Str3	-0.05	0.05	0.63	0.15	0.46	0.54	1.1
Str4	-0.02	0.06	0.57	-0.14	0.33	0.67	1.1
Str5	0.00	-0.10	0.59	0.06	0.34	0.66	1.1
Ang1	-0.07	0.11	0.37	0.37	0.35	0.65	2.2
Ang2	-0.02	0.08	-0.02	0.70	0.51	0.49	1.0
Ang3	0.11	0.10	0.06	0.59	0.43	0.57	1.2

	TC2	TC1	TC3	TC4
SS loadings	3.02	2.92	1.92	1.57
Proportion Var	0.12	0.12	0.08	0.06
Cumulative Var	0.12	0.24	0.31	0.38
Proportion Explained	0.32	0.31	0.20	0.17
Cumulative Proportion	0.32	0.63	0.83	1.00

With component correlations of

TC2	TC1	TC3	TC4
TC2	1.00	0.34	0.22
TC1	0.34	1.00	0.24
TC3	0.22	0.24	1.00
TC4	0.10	0.20	0.16
			1.00

Mean item complexity = 1.4

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07

Fit based upon off diagonal values = 0.86

We can make it a little easier to interpret by removing all factor loadings below .30.

```
psych::print.psych(pcaOBL, cut = 0.3, sort = TRUE)
```

## Principal Components Analysis

```
Call: psych::principal(r = items, nfactors = 4, rotate = "oblimin")
```

Standardized loadings (pattern matrix) based upon correlation matrix

item	TC2	TC1	TC3	TC4	h2	u2	com	
Obj8	8	0.67			0.43	0.57	1.1	
Obj2	2	0.61			0.44	0.56	1.3	
Obj1	1	0.57			0.37	0.63	1.2	
Obj10	10	0.55			0.33	0.67	1.3	
Obj4	4	0.54			0.36	0.64	1.2	
Obj6	6	0.45			0.25	0.75	1.7	
Obj7	7	0.44			0.36	0.64	1.6	
Obj9	9	0.43			0.23	0.77	1.5	
Obj3	3	0.43			0.32	0.68	1.7	
Obj5	5	0.39	0.33		0.35	0.65	2.3	
Marg1	11		0.69		0.52	0.48	1.1	
Marg7	17		0.65		0.38	0.62	1.1	
Marg5	15		0.64		0.42	0.58	1.1	
Marg3	13		0.60		0.39	0.61	1.0	
Marg2	12		0.58		0.41	0.59	1.2	
Marg4	14		0.55		0.40	0.60	1.3	
Marg6	16		0.45		0.38	0.62	1.8	
Str3	20			0.63	0.46	0.54	1.1	
Str5	22			0.59	0.34	0.66	1.1	
Str4	21			0.57	0.33	0.67	1.1	
Str1	18			0.53	0.40	0.60	1.6	
Str2	19			0.32	0.29	0.71	3.1	
Ang2	24				0.70	0.51	0.49	1.0
Ang3	25				0.59	0.43	0.57	1.2
Ang1	23			0.37	0.37	0.35	0.65	2.2

	TC2	TC1	TC3	TC4
SS loadings	3.02	2.92	1.92	1.57
Proportion Var	0.12	0.12	0.08	0.06
Cumulative Var	0.12	0.24	0.31	0.38
Proportion Explained	0.32	0.31	0.20	0.17
Cumulative Proportion	0.32	0.63	0.83	1.00

With component correlations of

	TC2	TC1	TC3	TC4
TC2	1.00	0.34	0.22	0.10
TC1	0.34	1.00	0.24	0.20
TC3	0.22	0.24	1.00	0.16
TC4	0.10	0.20	0.16	1.00

Mean item complexity = 1.4

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07

Fit based upon off diagonal values = 0.86

The oblique rotation perfectly replicated the GRMS solution from the Lewis and Neville [2015] article. Note, though, that because our specification included “sort=TRUE” that the relative weights wiggled around and so the items are listed in a different order than in the orthogonal rotation.

Let's create a table and write it to a file.

```
pcaOBL_table <- round(pcaOBL$loadings, 3)
write.table(pcaOBL_table, file = "pcaOBL_table.csv", sep = ",", col.names = TRUE,
            row.names = FALSE)
pcaOBL_table
```

## Loadings:

	TC2	TC1	TC3	TC4
Obj1	0.572			0.127
Obj2	0.611			-0.212
Obj3	0.428		0.273	
Obj4	0.536	0.158		
Obj5	0.391	0.327		-0.151
Obj6	0.446		-0.151	0.217
Obj7	0.443	0.103	0.168	0.147
Obj8	0.669		-0.114	
Obj9	0.435			-0.194
Obj10	0.549	-0.122		0.188
Marg1		0.694		0.127
Marg2		0.577	0.182	
Marg3		0.596		
Marg4	0.116	0.550		0.150
Marg5		0.640		-0.134
Marg6	0.190	0.446		0.208
Marg7		0.647		

Str1	0.191	0.532	-0.208		
Str2	0.243	0.320	0.294		
Str3		0.635	0.145		
Str4		0.568	-0.137		
Str5		-0.100	0.586		
Ang1		0.107	0.368		
Ang2			0.700		
Ang3	0.110	0.101	0.589		
		TC2	TC1	TC3	TC4
SS loadings		2.832	2.731	1.809	1.488
Proportion Var	0.113	0.109	0.072	0.060	
Cumulative Var	0.113	0.223	0.295	0.354	

The same four components/scales have emerged, but they are in different order.

The oblique rotation allows us to see the correlation between the components/scales. This was not available in the orthogonal rotation because the assumption of the orthogonal/varimax rotation is that the scales/components are uncorrelated; hence in the analysis they were fixed to 0.0.

We can see that all the scales have low to moderate (i.e, 0.10 to 0.34) correlations with each other.

Of course, there is always a little complexity. In oblique rotations, there is a distinction between the *pattern* matrix (which reports component loadings and is comparable to the matrix we interpreted for the orthogonal rotation) and the *structure* matrix (takes into account the relationship between the components/scales – it is a product of the pattern matrix and the matrix containing the correlation coefficients between the components/scales). Most interpret the pattern matrix because it is simpler; however, it could be that values in the pattern matrix are suppressed because of relations between the components. Therefore, the structure matrix can be a useful check and some editors will request it.

Obtaining the structure matrix requires two steps. First, we multiply the factor loadings with the phi matrix.

```
# names(pcaOBL)
pcaOBL$loadings %*% pcaOBL$Phi
```

	TC2	TC1	TC3	TC4
Obj1	0.58824537	0.1950747	0.21661519	0.190694573
Obj2	0.62785704	0.2390937	0.21155257	-0.122623113
Obj3	0.49675336	0.2318812	0.37521696	0.108241426
Obj4	0.57498964	0.3258365	0.08444990	0.098254346
Obj5	0.49580117	0.4393738	0.18560641	-0.037041960
Obj6	0.43509862	0.1601298	-0.01736119	0.239798447
Obj7	0.53068509	0.3216191	0.31406071	0.240511062
Obj8	0.64219716	0.1974906	0.03287727	0.034233149
Obj9	0.44294690	0.1882983	0.09315759	-0.132090129
Obj10	0.53544723	0.1098069	0.15864543	0.225669435
Marg1	0.20399037	0.7081844	0.19534185	0.264854138

	TC2	TC1	TC3	TC4
Marg2	0.17453400	0.6091293	0.31218266	0.195725778
Marg3	0.25184667	0.6242636	0.17230976	0.179169214
Marg4	0.29865870	0.5994737	0.09538494	0.259471246
Marg5	0.23165634	0.6308766	0.18374798	0.004895817
Marg6	0.36039152	0.5497905	0.16994812	0.315569585
Marg7	0.17404434	0.6022982	0.04196051	0.024371153
Str1	0.31852390	0.2385467	0.56279443	-0.085382899
Str2	0.31163120	0.1189926	0.39749589	0.349829371
Str3	0.12040010	0.2143818	0.65842331	0.251075007
Str4	0.10844560	0.1581938	0.55545185	-0.037207616
Str5	0.10520418	0.0511699	0.57214012	0.131316282
Ang1	0.08774735	0.2451910	0.43622528	0.443610249
Ang2	0.07752669	0.2150177	0.10924533	0.712602474
Ang3	0.21857516	0.2707787	0.20074776	0.629758227

Next, we can use Field's [2012] function to produce the matrix.

```
# Field's function to produce the structure matrix
factor.structure <- function(fa, cut = 0.2, decimals = 2) {
  structure.matrix <- psych::fa.sort(fa$loadings %*% fa$Phi)
  structure.matrix <- data.frame(ifelse(abs(structure.matrix) < cut,
                                         "", round(structure.matrix, decimals)))
  return(structure.matrix)
}

factor.structure(pca0BL, cut = 0.3)
```

	TC2	TC1	TC3	TC4
Obj8	0.64			
Obj2	0.63			
Obj1	0.59			
Obj4	0.57	0.33		
Obj10	0.54			
Obj7	0.53	0.32	0.31	
Obj3	0.5		0.38	
Obj5	0.5	0.44		
Obj9	0.44			
Obj6	0.44			
Marg1		0.71		
Marg5		0.63		
Marg3		0.62		
Marg2		0.61	0.31	
Marg7		0.6		
Marg4		0.6		
Marg6	0.36	0.55		0.32
Str3			0.66	
Str5			0.57	

```

Str1  0.32      0.56
Str4           0.56
Str2  0.31      0.4  0.35
Ang2           0.71
Ang3           0.63
Ang1           0.44  0.44

```

Although some of the relative values changed, our items were stable regarding their component membership.

### 8.5.5 Component Scores

Component *scores* (PC scores) can be created for each case (row) on each component (column). These can be used to assess the relative standing of one person on the construct/variable to another. We can also use them in regression (in place of means or sums) when groups of predictors correlate so highly that there is multicollinearity.

Computation involves multiplying an individual's item-level responses by the component loadings we obtained through the PCA process. The results will be one score per component for each row/case.

```
pcaOBL <- psych::principal(items, nfactors = 4, rotate = "oblimin", scores = TRUE)
head(pcaOBL$scores, 10)  #shows us only the first 10 (of N = 2571)
```

	TC2	TC1	TC3	TC4
[1,]	-0.63652243	-0.4689396	-0.2566274	-0.93492046
[2,]	0.44263027	-1.3441810	0.7781981	0.10235218
[3,]	0.60445992	0.5001073	0.2127395	0.55484479
[4,]	-0.77336136	-1.1297799	0.5455242	-1.39690887
[5,]	-0.65462849	1.4645126	-1.9729147	-0.01302103
[6,]	-0.33309990	1.0565678	-0.1327958	0.68245164
[7,]	0.41890349	0.4058683	-0.8434724	-1.21634748
[8,]	-1.46968183	-1.1418916	0.3745886	2.19465772
[9,]	-0.76230611	-1.3266916	-0.8275204	0.83138284
[10,]	-0.04501587	-0.1844763	-0.8872830	0.79893269

```
dfGRMS <- cbind(items, pcaOBL$scores)  #adds them to our raw dataset
```

To bring this full circle, we can see the correlation of the component scores; the pattern maps onto what we saw previously.

```
psych::corr.test(dfGRMS[c("TC1", "TC4", "TC3", "TC2")])
```

```

Call:psych::corr.test(x = dfGRMS[c("TC1", "TC4", "TC3", "TC2")])
Correlation matrix
  TC1  TC4  TC3  TC2

```

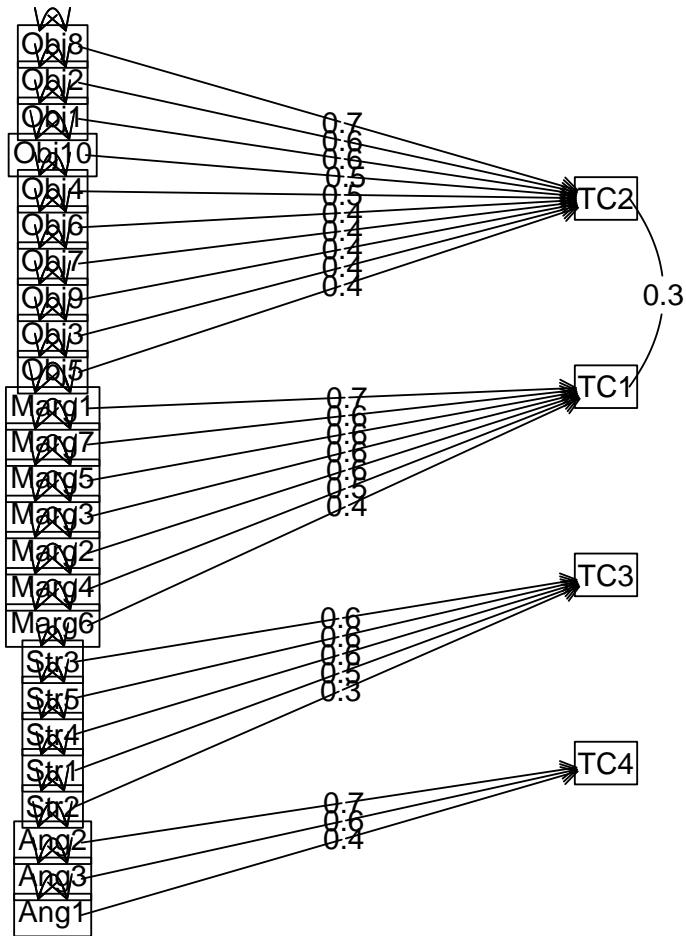
```
TC1 1.00 0.20 0.24 0.34
TC4 0.20 1.00 0.16 0.10
TC3 0.24 0.16 1.00 0.22
TC2 0.34 0.10 0.22 1.00
Sample Size
[1] 259
Probability values (Entries above the diagonal are adjusted for multiple tests.)
   TC1  TC4  TC3  TC2
TC1  0 0.00 0.00 0.0
TC4  0 0.00 0.02 0.1
TC3  0 0.01 0.00 0.0
TC2  0 0.10 0.00 0.0
```

To see confidence intervals of the correlations, print with the `short=FALSE` option

And now for a figure of the oblique rotation. Note that figure includes semi-circles between TC1/TC2 and TC1/TC4. These represent significant correlation coefficients between the components that are named. In contrast, the orthogonal rotation required the components to be uncorrelated.

```
psych::fa.diagram(pcaOBL, error = TRUE, side = 3)
```

## Components Analysis



## 8.6 APA Style Results

### Results

The dimensionality of the 25 items from the Gendered Racial Microaggressions Scale for Black Women was analyzed using principal components analysis. Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was .84, indicating acceptable sampling adequacy. The Bartlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters

of correlated variables. In our dataset,  $\chi^2(300) = 1113.299, p < .001$ , indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.01140, supporting the suitability of our data for analysis.

Four criteria were used to determine the number of components to extract: a priori theory, the scree test, the eigenvalue-greater-than-one criteria, and the interpretability of the solution. Kaiser's eigenvalue-greater-than-one criteria suggested seven components, and, in combination explained 38% of the variance. The inflection in the scree plot suggested retaining between one and four components. Considering the a priori theory obtained from the original psychometric article [Lewis and Neville, 2015], four components were extracted. We investigated each with orthogonal (varimax) and oblique (oblimin) procedures. Given the low-to-moderate correlations (ranging from 0.10 to 0.30) and the clear component loadings, we determined that an oblique solution was most appropriate.

The rotated solution, as shown in Table 1 and Figure 1, yielded four interpretable components, each listed with the proportion of variance accounted for: assumptions of beauty and sexual objectification (12%), silenced and marginalized (12%), strong woman stereotype (8%), and angry woman stereotype (6%).

Regarding the Table 1, I would include a table with all the values, bolding those with component membership. This is easy, though, because it is how the table was exported when we wrote it to a .csv file.

## 8.7 Back to the FutuRe: The relationship between PCA and item analysis

I included the lesson on item analysis because I find it to be a useful stepping stone into principal components and principal factor analyses. How do the results we obtained from PCA compare to those found in item analysis?

First, we score the total and subscales using the dataset we simulated above (dfGRMS).

```
library(tidyverse)
GRMSVars <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6", "Obj7", "Obj8",
               "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4", "Marg5", "Marg6",
               "Marg7", "Str1", "Str2", "Str3", "Str4", "Str5", "Ang1", "Ang2", "Ang3")
ObjectifiedVars <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6", "Obj7",
                      "Obj8", "Obj9", "Obj10")
MarginalizedVars <- c("Marg1", "Marg2", "Marg3", "Marg4", "Marg5", "Marg6",
                      "Marg7")
StrongVars <- c("Str1", "Str2", "Str3", "Str4", "Str5")
AngryVars <- c("Ang1", "Ang2", "Ang3")
```

```
items$GRMStot <- sjstats::mean_n(items[, GRMSVars], 0.8) #will create the mean for each indiv
items$Objectified <- sjstats::mean_n(items[, ObjectifiedVars], 0.8) #will create the mean for
items$Marginalized <- sjstats::mean_n(items[, MarginalizedVars], 0.8) #will create the mean f
items$Strong <- sjstats::mean_n(items[, StrongVars], 0.8) #will create the mean for each indiv
items$Angry <- sjstats::mean_n(items[, AngryVars], 0.8) #will create the mean for each indivi
```

While we are at it, let's just create tiny dfs with just our variables of interest.

```
GRMStotal <- dplyr::select(items, Obj1:Ang3)
Objectification <- dplyr::select(items, Obj1:Obj10)
Marginalization <- dplyr::select(items, Marg1:Marg7)
Strong <- dplyr::select(items, Str1:Str5)
Angry <- dplyr::select(items, Ang1:Ang3)
```

## 8.7.1 Calculating and Extracting Item-Total Correlation Coefficients

### 8.7.1.1 Corrected item-total correlations from the *psych::alpha()*

Let's first ask, "Is there support for this instrument as a unidimensional measure?" To do that, we get an alpha for the whole scale score.

```
GRMSalpha <- psych::alpha(GRMStotal) #creating an object from this analysis so I can extract
GRMSalpha
```

Reliability analysis

Call: psych::alpha(x = GRMStotal)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.83	0.83	0.85	0.16	4.8	0.015	3	0.34	0.16

95% confidence boundaries

lower alpha upper

Feldt 0.8 0.83 0.86

Duhachek 0.8 0.83 0.86

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Obj1	0.82	0.82	0.84	0.16	4.6	0.016	0.0063	0.16	
Obj2	0.82	0.82	0.84	0.16	4.6	0.016	0.0063	0.16	
Obj3	0.82	0.82	0.84	0.16	4.6	0.016	0.0066	0.15	
Obj4	0.82	0.82	0.84	0.16	4.6	0.016	0.0063	0.16	
Obj5	0.82	0.82	0.84	0.16	4.6	0.016	0.0063	0.16	
Obj6	0.83	0.83	0.84	0.17	4.8	0.016	0.0065	0.17	
Obj7	0.82	0.82	0.83	0.16	4.5	0.016	0.0063	0.15	
Obj8	0.82	0.82	0.84	0.16	4.7	0.016	0.0062	0.16	

Obj9	0.83	0.83	0.84	0.17	4.8	0.015	0.0062	0.17
Obj10	0.82	0.82	0.84	0.16	4.7	0.016	0.0064	0.16
Marg1	0.82	0.82	0.83	0.16	4.5	0.016	0.0057	0.16
Marg2	0.82	0.82	0.84	0.16	4.6	0.016	0.0062	0.16
Marg3	0.82	0.82	0.84	0.16	4.6	0.016	0.0061	0.16
Marg4	0.82	0.82	0.84	0.16	4.6	0.016	0.0061	0.16
Marg5	0.82	0.82	0.84	0.16	4.6	0.016	0.0062	0.16
Marg6	0.82	0.82	0.83	0.16	4.5	0.016	0.0062	0.16
Marg7	0.82	0.82	0.84	0.16	4.7	0.016	0.0062	0.16
Str1	0.82	0.82	0.84	0.16	4.7	0.016	0.0066	0.16
Str2	0.82	0.82	0.84	0.16	4.7	0.016	0.0065	0.16
Str3	0.82	0.82	0.84	0.16	4.7	0.016	0.0065	0.16
Str4	0.83	0.83	0.84	0.17	4.8	0.015	0.0062	0.17
Str5	0.83	0.83	0.84	0.17	4.8	0.015	0.0062	0.17
Ang1	0.82	0.83	0.84	0.16	4.7	0.016	0.0064	0.16
Ang2	0.83	0.83	0.84	0.17	4.8	0.015	0.0063	0.17
Ang3	0.82	0.82	0.84	0.16	4.7	0.016	0.0065	0.16

## Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd	
Obj1	259	0.48	0.49	0.46	0.41	3.2	0.71	
Obj2	259	0.47	0.48	0.45	0.40	3.1	0.75	
Obj3	259	0.49	0.49	0.46	0.42	3.1	0.78	
Obj4	259	0.50	0.49	0.46	0.42	3.1	0.82	
Obj5	259	0.50	0.50	0.47	0.43	2.9	0.77	
Obj6	259	0.36	0.37	0.31	0.28	2.9	0.76	
Obj7	259	0.57	0.55	0.53	0.48	3.0	0.99	
Obj8	259	0.43	0.44	0.41	0.36	3.0	0.70	
Obj9	259	0.34	0.34	0.29	0.26	3.1	0.77	
Obj10	259	0.41	0.42	0.38	0.34	2.8	0.69	
Marg1	259	0.53	0.53	0.52	0.46	3.2	0.75	
Marg2	259	0.50	0.50	0.48	0.43	3.2	0.72	
Marg3	259	0.51	0.50	0.48	0.42	2.9	0.88	
Marg4	259	0.50	0.51	0.48	0.43	3.0	0.71	
Marg5	259	0.47	0.47	0.44	0.40	2.9	0.72	
Marg6	259	0.55	0.55	0.53	0.48	3.1	0.78	
Marg7	259	0.41	0.40	0.36	0.32	3.0	0.83	
Str1	259	0.42	0.43	0.39	0.35	3.2	0.73	
Str2	259	0.41	0.41	0.37	0.33	3.1	0.74	
Str3	259	0.42	0.42	0.38	0.34	2.8	0.81	
Str4	259	0.30	0.31	0.25	0.23	2.9	0.68	
Str5	259	0.31	0.30	0.24	0.22	3.1	0.83	
Ang1	259	0.41	0.39	0.35	0.32	3.0	0.86	
Ang2	259	0.35	0.34	0.29	0.26	2.8	0.83	
Ang3	259	0.43	0.43	0.39	0.36	2.9	0.76	

Non missing response frequency for each item

1    2    3    4    5 miss

Obj1	0.01	0.11	0.52	0.34	0.02	0
Obj2	0.01	0.19	0.56	0.22	0.03	0
Obj3	0.02	0.19	0.53	0.24	0.02	0
Obj4	0.01	0.20	0.50	0.23	0.06	0
Obj5	0.02	0.26	0.50	0.20	0.01	0
Obj6	0.01	0.30	0.50	0.17	0.02	0
Obj7	0.07	0.24	0.41	0.23	0.06	0
Obj8	0.02	0.19	0.59	0.20	0.01	0
Obj9	0.01	0.19	0.49	0.29	0.02	0
Obj10	0.02	0.30	0.56	0.12	0.00	0
Marg1	0.01	0.13	0.53	0.29	0.03	0
Marg2	0.00	0.15	0.53	0.30	0.02	0
Marg3	0.05	0.25	0.45	0.22	0.02	0
Marg4	0.01	0.20	0.57	0.20	0.01	0
Marg5	0.01	0.29	0.54	0.15	0.01	0
Marg6	0.01	0.19	0.47	0.31	0.02	0
Marg7	0.03	0.22	0.49	0.23	0.03	0
Str1	0.01	0.12	0.56	0.27	0.03	0
Str2	0.02	0.17	0.56	0.23	0.02	0
Str3	0.03	0.33	0.44	0.19	0.02	0
Str4	0.01	0.27	0.58	0.14	0.01	0
Str5	0.02	0.23	0.42	0.32	0.02	0
Ang1	0.02	0.26	0.42	0.27	0.03	0
Ang2	0.03	0.34	0.45	0.15	0.03	0
Ang3	0.03	0.24	0.54	0.17	0.01	0

And now each of the subscales:

```
ObjAlpha <- psych::alpha(Objetification) #creating an object from this analysis so I can ext...
```

Reliability analysis  
Call: psych::alpha(x = Objetification)

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.74	0.74	0.73	0.22	2.9	0.024	3	0.43	0.23

95% confidence boundaries  
lower alpha upper

Feldt	0.69	0.74	0.78
Duhachek	0.69	0.74	0.79

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Obj1	0.71	0.72	0.70	0.22	2.5	0.026	0.0034	0.22	
Obj2	0.71	0.71	0.69	0.21	2.4	0.027	0.0036	0.21	

Obj3	0.72	0.72	0.70	0.22	2.6	0.026	0.0039	0.23
Obj4	0.71	0.71	0.70	0.22	2.5	0.027	0.0035	0.22
Obj5	0.72	0.72	0.70	0.22	2.6	0.026	0.0032	0.23
Obj6	0.73	0.74	0.72	0.24	2.8	0.025	0.0031	0.25
Obj7	0.72	0.72	0.70	0.22	2.5	0.026	0.0036	0.22
Obj8	0.71	0.71	0.70	0.22	2.5	0.027	0.0033	0.21
Obj9	0.73	0.73	0.72	0.24	2.8	0.025	0.0030	0.24
Obj10	0.72	0.72	0.71	0.23	2.6	0.026	0.0035	0.22

## Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd	
Obj1	259	0.56	0.57	0.50	0.42	3.2	0.71	
Obj2	259	0.60	0.61	0.55	0.47	3.1	0.75	
Obj3	259	0.55	0.55	0.47	0.41	3.1	0.78	
Obj4	259	0.60	0.59	0.52	0.45	3.1	0.82	
Obj5	259	0.55	0.55	0.47	0.40	2.9	0.77	
Obj6	259	0.45	0.46	0.34	0.29	2.9	0.76	
Obj7	259	0.61	0.57	0.50	0.43	3.0	0.99	
Obj8	259	0.59	0.60	0.53	0.46	3.0	0.70	
Obj9	259	0.47	0.47	0.36	0.31	3.1	0.77	
Obj10	259	0.51	0.52	0.43	0.37	2.8	0.69	

## Non missing response frequency for each item

	1	2	3	4	5	miss
Obj1	0.01	0.11	0.52	0.34	0.02	0
Obj2	0.01	0.19	0.56	0.22	0.03	0
Obj3	0.02	0.19	0.53	0.24	0.02	0
Obj4	0.01	0.20	0.50	0.23	0.06	0
Obj5	0.02	0.26	0.50	0.20	0.01	0
Obj6	0.01	0.30	0.50	0.17	0.02	0
Obj7	0.07	0.24	0.41	0.23	0.06	0
Obj8	0.02	0.19	0.59	0.20	0.01	0
Obj9	0.01	0.19	0.49	0.29	0.02	0
Obj10	0.02	0.30	0.56	0.12	0.00	0

```
MargAlpha <- psych::alpha(Marginalization) #creating an object from this analysis so I can ex
MargAlpha
```

## Reliability analysis

```
Call: psych::alpha(x = Marginalization)
```

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.74	0.75	0.72		0.3	3	0.024	3	0.49
								0.28

95% confidence boundaries  
lower alpha upper

```
Feldt    0.69  0.74  0.79
Duhachek 0.70  0.74  0.79
```

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Marg1	0.70	0.70	0.66	0.28	2.3	0.029	0.0021	0.27	
Marg2	0.71	0.72	0.68	0.30	2.5	0.027	0.0032	0.28	
Marg3	0.71	0.71	0.68	0.29	2.5	0.028	0.0028	0.28	
Marg4	0.71	0.72	0.68	0.30	2.5	0.027	0.0018	0.29	
Marg5	0.72	0.72	0.69	0.30	2.6	0.027	0.0034	0.29	
Marg6	0.72	0.72	0.69	0.30	2.6	0.027	0.0039	0.29	
Marg7	0.73	0.73	0.70	0.31	2.7	0.026	0.0031	0.31	

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
Marg1	259	0.69	0.70	0.64	0.54	3.2	0.75
Marg2	259	0.62	0.63	0.53	0.46	3.2	0.72
Marg3	259	0.67	0.65	0.57	0.49	2.9	0.88
Marg4	259	0.62	0.63	0.54	0.46	3.0	0.71
Marg5	259	0.59	0.61	0.50	0.43	2.9	0.72
Marg6	259	0.63	0.62	0.52	0.45	3.1	0.78
Marg7	259	0.59	0.58	0.46	0.39	3.0	0.83

Non missing response frequency for each item

	1	2	3	4	5	miss
Marg1	0.01	0.13	0.53	0.29	0.03	0
Marg2	0.00	0.15	0.53	0.30	0.02	0
Marg3	0.05	0.25	0.45	0.22	0.02	0
Marg4	0.01	0.20	0.57	0.20	0.01	0
Marg5	0.01	0.29	0.54	0.15	0.01	0
Marg6	0.01	0.19	0.47	0.31	0.02	0
Marg7	0.03	0.22	0.49	0.23	0.03	0

```
StrongAlpha <- psych::alpha(Strong) #creating an object from this analysis so I can extract a
```

Reliability analysis

Call: psych::alpha(x = Strong)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.52	0.52	0.47	0.18	1.1	0.047	3	0.45	0.18

95% confidence boundaries

	lower	alpha	upper
Feldt	0.43	0.52	0.61
Duhachek	0.43	0.52	0.62

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Str1	0.47	0.48	0.41	0.19	0.91	0.053	0.00247	0.18	
Str2	0.47	0.48	0.41	0.19	0.91	0.053	0.00304	0.18	
Str3	0.42	0.43	0.36	0.16	0.74	0.058	0.00060	0.16	
Str4	0.48	0.48	0.41	0.19	0.91	0.053	0.00049	0.18	
Str5	0.49	0.49	0.42	0.19	0.96	0.051	0.00209	0.18	

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
Str1	259	0.57	0.58	0.39	0.28	3.2	0.73
Str2	259	0.57	0.58	0.38	0.28	3.1	0.74
Str3	259	0.65	0.64	0.50	0.36	2.8	0.81
Str4	259	0.55	0.58	0.39	0.28	2.9	0.68
Str5	259	0.59	0.56	0.36	0.26	3.1	0.83

Non missing response frequency for each item

	1	2	3	4	5	miss
Str1	0.01	0.12	0.56	0.27	0.03	0
Str2	0.02	0.17	0.56	0.23	0.02	0
Str3	0.03	0.33	0.44	0.19	0.02	0
Str4	0.01	0.27	0.58	0.14	0.01	0
Str5	0.02	0.23	0.42	0.32	0.02	0

```
AngryAlpha <- psych::alpha(Angry) #creating an object from this analysis so I can extract and
AngryAlpha
```

Reliability analysis

Call: psych::alpha(x = Angry)

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
	0.5	0.5	0.4	0.25	1	0.054	2.9	0.58	0.26

95% confidence boundaries

	lower	alpha	upper
Feldt	0.39	0.5	0.60
Duhachek	0.40	0.5	0.61

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Ang1	0.42	0.42	0.27	0.27	0.72	0.072	NA	0.27	
Ang2	0.42	0.42	0.26	0.26	0.72	0.072	NA	0.26	
Ang3	0.37	0.37	0.23	0.23	0.59	0.078	NA	0.23	

Item statistics

```

      n raw.r std.r r.cor r.drop mean   sd
Ang1 259  0.72  0.70  0.44  0.31  3.0 0.86
Ang2 259  0.71  0.70  0.44  0.31  2.8 0.83
Ang3 259  0.69  0.72  0.48  0.34  2.9 0.76

Non missing response frequency for each item
      1    2    3    4    5 miss
Ang1 0.02 0.26 0.42 0.27 0.03    0
Ang2 0.03 0.34 0.45 0.15 0.03    0
Ang3 0.03 0.24 0.54 0.17 0.01    0

```

### 8.7.1.2 Correlating items with other subscale totals

```
Obj_othR <- psych::corr.test(items[c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5",
  "Obj6", "Obj7", "Obj8", "Obj9", "Obj10", "Marginalized", "Strong",
  "Angry")])
```

```
Marg_othR <- psych::corr.test(items[c("Marg1", "Marg2", "Marg3", "Marg4",
  "Marg5", "Marg6", "Marg7", "Objectified", "Strong", "Angry")])
```

```
Str_othR <- psych::corr.test(items[c("Str1", "Str2", "Str3", "Str4", "Str5",
  "Objectified", "Marginalized", "Angry")])
```

```
Ang_othR <- psych::corr.test(items[c("Ang1", "Ang2", "Ang3", "Objectified",
  "Marginalized", "Strong")])
```

### 8.7.1.3 Extracting values, binding them together, and joining the files

```

# names(Obj_other) Extracting the item-level statistics from the
# alpha object
Obj_othR <- as.data.frame(Obj_othR$r) #Makes the item-total(other) correlation matrix a df
# Adding variable names so we don't get lost
Obj_othR$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6", "Obj7",
  "Obj8", "Obj9", "Obj10", "Marginalized", "Strong", "Angry")
# deleting the rows with the total scale scores
Obj_othR <- Obj_othR[!Obj_othR$Items == "Marginalized", ]
Obj_othR <- Obj_othR[!Obj_othR$Items == "Strong", ]
Obj_othR <- Obj_othR[!Obj_othR$Items == "Angry", ]
Obj_othR[, "Objectified"] <- NA #We need a column for this to bind the items, later.
Obj_othR <- dplyr::select(Obj_othR, Items, Objectified, Marginalized, Strong,
  Angry) #Putting items in order
# Item Corrected Total Correlations
ObjAlpha <- as.data.frame(ObjAlpha$item.stats) #Grabbing the alpha objet we created earlier at

```

```

ObjAlpha$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6", "Obj7",
  "Obj8", "Obj9", "Obj10")
# Joining the two and selecting the vars of interest
ObjStats <- full_join(ObjAlpha, Obj_othR, by = "Items")
ObjStats$Objectified <- ObjStats$r.drop #Copy the item-corrected total (r.drop) into the Obj
ObjStats <- dplyr::select(ObjStats, Items, Objectified, Marginalized, Strong,
  Angry)
# rm(ObjAlpha, Obj_othR) #It's messay, dropping all the
# no-longer-necessary objects from the Global Environment

# Extracting the item-level statistics from the alpha object
Marg_othR <- as.data.frame(Marg_othR$r) #Makes the item-total(other) correlation matrix a df
# Adding variable names so we don't get lost
Marg_othR$Items <- c("Marg1", "Marg2", "Marg3", "Marg4", "Marg5", "Marg6",
  "Marg7", "Objectified", "Strong", "Angry")
# deleting the rows with the total scale scores
Marg_othR <- Marg_othR[!Marg_othR$Items == "Objectified", ]
Marg_othR <- Marg_othR[!Marg_othR$Items == "Strong", ]
Marg_othR <- Marg_othR[!Marg_othR$Items == "Angry", ]
Marg_othR[, "Marginalized"] <- NA #We need a column for this to bind the items, later.
Marg_othR <- dplyr::select(Marg_othR, Items, Objectified, Marginalized,
  Strong, Angry)
# Item Corrected Total Correlations
MargAlpha <- as.data.frame(MargAlpha$item.stats) #Grabbing the alpha objet we created earlier
MargAlpha$Items <- c("Marg1", "Marg2", "Marg3", "Marg4", "Marg5", "Marg6",
  "Marg7")
# Joining the two and selecting the vars of interest
MargStats <- full_join(MargAlpha, Marg_othR, by = "Items")
MargStats$Marginalized <- MargStats$r.drop #Copy the item-corrected total (r.drop) into the M
MargStats <- dplyr::select(MargStats, Items, Objectified, Marginalized,
  Strong, Angry)
# rm(MargAlpha, Marg_othR) #It's messay, dropping all the
# no-longer-necessary objects from the Global Environment

Str_othR <- as.data.frame(Str_othR$r) #Makes the item-total(other) correlation matrix a df
# Adding variable names so we don't get lost
Str_othR$Items <- c("Strong1", "Strong2", "Strong3", "Strong4", "Strong5",
  "Objectified", "Marginalized", "Angry")
# deleting the rows with the total scale scores
Str_othR <- Str_othR[!Str_othR$Items == "Objectified", ]
Str_othR <- Str_othR[!Str_othR$Items == "Marginalized", ]
Str_othR <- Str_othR[!Str_othR$Items == "Angry", ]
Str_othR[, "Strong"] <- NA
Str_othR <- dplyr::select(Str_othR, Items, Objectified, Marginalized, Strong,
  Angry)
# Item Corrected Total Correlations

```

```

StrongAlpha <- as.data.frame(StrongAlpha$item.stats) #Grabbing the alpha objet we created earlier
StrongAlpha$Items <- c("Strong1", "Strong2", "Strong3", "Strong4", "Strong5")
# Joining the two and selecting the vars of interest
StrStats <- full_join(StrongAlpha, Str_othR, by = "Items")
StrStats$Strong <- StrStats$r.drop #Copy the item-corrected total (r.drop) into the Strong variable
StrStats <- dplyr::select(StrStats, Items, Objectified, Marginalized, Strong,
                         Angry)
rm(StrongAlpha, Str_othR) #It's message, dropping all the no-longer-necessary objects from the environment

Ang_othR <- as.data.frame(Ang_othR$r) #Makes the item-total(other) correlation matrix a df
# Adding variable names so we don't get lost
Ang_othR$Items <- c("Angry1", "Angry2", "Angry3", "Objectified", "Marginalized",
                   "Strong")
# deleting the rows with the total scale scores
Ang_othR <- Ang_othR[!Ang_othR$Items == "Objectified", ]
Ang_othR <- Ang_othR[!Ang_othR$Items == "Marginalized", ]
Ang_othR <- Ang_othR[!Ang_othR$Items == "Strong", ]
Ang_othR[, "Angry"] <- NA
Ang_othR <- dplyr::select(Ang_othR, Items, Objectified, Marginalized, Strong,
                           Angry)
# Item Corrected Total Correlations
AngryAlpha <- as.data.frame(AngryAlpha$item.stats) #Grabbing the alpha objet we created earlier
AngryAlpha$Items <- c("Angry1", "Angry2", "Angry3")
# Joining the two and selecting the vars of interest
AngStats <- full_join(AngryAlpha, Ang_othR, by = "Items")
AngStats$Angry <- AngStats$r.drop #Copy the item-corrected total (r.drop) into the Angry variable
AngStats <- dplyr::select(AngStats, Items, Objectified, Marginalized, Strong,
                         Angry)
rm(AngryAlpha, Ang_othR) #It's message, dropping all the no-longer-necessary objects from the environment

# Adding all the variables into a single table
ItemAnalysis <- rbind(ObjStats, MargStats, StrStats, AngStats)

# Preparing and adding the r.drop for total scale score
TotAlpha <- as.data.frame(GRMSalpha$item.stats)
TotAlpha$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6", "Obj7",
                    "Obj8", "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4", "Marg5",
                    "Marg6", "Marg7", "Strong1", "Strong2", "Strong3", "Strong4", "Strong5",
                    "Angry1", "Angry2", "Angry3")
TotAlpha <- dplyr::select(TotAlpha, Items, r.drop) #deleting the rows with the total scale scores

# Adding the r.drop for the total scale score
ItemAnalysis <- full_join(TotAlpha, ItemAnalysis, by = "Items")

# Adding the values from the Orthogonal rotation
pcaORTH_loadings <- data.frame(unclass(pcaORTH$loadings)) #I had to add 'unclass' to the loadings

```

```

pcaORTH_loadings$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6",
  "Obj7", "Obj8", "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4",
  "Marg5", "Marg6", "Marg7", "Strong1", "Strong2", "Strong3", "Strong4",
  "Strong5", "Angry1", "Angry2", "Angry3") #Item names for joining (and to make sure we know
# Deleting those lower rows pcaORTH_loadings <-
# pcaORTH_loadings[!pcaORTH_loadings$Items == 'GRMSTot',]
# pcaORTH_loadings <- pcaORTH_loadings[!pcaORTH_loadings$Items ==
# 'Objectified',] pcaORTH_loadings <-
# pcaORTH_loadings[!pcaORTH_loadings$Items == 'Marginalized',]
# pcaORTH_loadings <- pcaORTH_loadings[!pcaORTH_loadings$Items ==
# 'Strong',] pcaORTH_loadings <-
# pcaORTH_loadings[!pcaORTH_loadings$Items == 'Angry',]
pcaORTH_loadings <- rename(pcaORTH_loadings, objORTH = RC1, margORTH = RC2,
  strORTH = RC3, angORTH2 = RC4)

# Joining with the Item Stats
Comparisons <- full_join(ItemAnalysis, pcaORTH_loadings, by = "Items") #I had to add 'unclass'

# Adding the oblique loadings
pcaOBLQ_loadings <- data.frame(unclass(pcaOBL$loadings)) #I had to add 'unclass' to the loadings
pcaOBLQ_loadings$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6",
  "Obj7", "Obj8", "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4",
  "Marg5", "Marg6", "Marg7", "Strong1", "Strong2", "Strong3", "Strong4",
  "Strong5", "Angry1", "Angry2", "Angry3") #Item names for joining (and to make sure we know
# Deleting those lower rows pcaOBLQ_loadings <-
# pcaOBLQ_loadings[!pcaORTH_loadings$Items == 'GRMSTot',]
# pcaOBLQ_loadings <- pcaOBLQ_loadings[!pcaORTH_loadings$Items ==
# 'Objectified',] pcaOBLQ_loadings <-
# pcaOBLQ_loadings[!pcaORTH_loadings$Items == 'Marginalized',]
# pcaOBLQ_loadings <- pcaOBLQ_loadings[!pcaORTH_loadings$Items ==
# 'Strong',] pcaOBLQ_loadings <-
# pcaOBLQ_loadings[!pcaORTH_loadings$Items == 'Angry',]
pcaOBLQ_loadings <- rename(pcaOBLQ_loadings, objOBLQ = TC1, margOBLQ = TC2,
  strOBLQ = TC3, angOBLQ = TC4)

# Joining with the Item Stats
Comparisons <- full_join(Comparisons, pcaOBLQ_loadings, by = "Items") #I had to add 'unclass'

write.csv(Comparisons, file = "GRMS_Comparisons.csv", sep = ",", row.names = FALSE,
  col.names = TRUE) #Writes the table to a .csv file where you can open it with Excel and f

```

Warning in write.csv(Comparisons, file = "GRMS\_Comparisons.csv", sep = ",", :  
attempt to set 'col.names' ignored

Warning in write.csv(Comparisons, file = "GRMS\_Comparisons.csv", sep = ",", :  
attempt to set 'sep' ignored

```
saveRDS(Comparisons, "GRMS_Comparisons.rds") #Writes the file as an .rds so that if anything
```

#### 8.7.1.4 Interpreting the result

The result of this work is a table that includes:

- **r.drop** Corrected item-total (entire GRMS) coefficients
- **Item-total correlations** of the items correlated with their own subscale (bold; correlation does not include the item being correlated) and the other subscales
- **PCA: Orthogonal rotation** factor loadings of the four-scales with a rotation that maximizes the independents (uncorrelatedness) of the scales
- **PCA: Oblique rotation** factor loadings of the four-scales with a rotation that permits correlation between subscales

Items	Total Scale	Item Corrected-Total (bold) and Item-Other Scale Correlations				Principal Components Analysis: Orthogonal Rotation				Principal Components Analysis: Oblique Rotation				
		r.drop	Objectified	Marginalized	Strong	Angry	margORTH	objORTH	strORTH	angORTH2	margOBLQ	objOBLQ	strOBLQ	angOBLQ
				0.41	0.42	0.26	0.26	0.14	0.56	0.07	0.13	0.17	0.57	-0.04
Obj1		0.40	0.47	0.26	0.22	0.08	0.61	0.14	0.16	-0.15	0.61	0.05	0.10	-0.21
Obj2		0.42	0.41	0.26	0.29	0.19	0.45	0.12	0.31	0.08	0.43	0.02	0.27	0.02
Obj3		0.42	0.45	0.30	0.17	0.19	0.55	0.23	0.00	0.07	0.54	0.16	-0.08	0.02
Obj4		0.43	0.40	0.36	0.21	0.14	0.44	0.37	0.12	-0.08	0.39	0.33	0.05	-0.15
Obj5		0.28	0.29	0.20	0.13	0.14	0.43	0.08	-0.10	0.23	0.45	0.00	-0.15	0.22
Obj6		0.48	0.43	0.35	0.29	0.26	0.47	0.20	0.22	0.21	0.44	0.10	0.17	0.15
Obj7		0.36	0.46	0.23	0.13	0.09	0.64	0.09	-0.04	0.01	0.67	0.00	-0.11	-0.02
Obj8		0.26	0.31	0.17	0.09	0.09	0.44	0.13	0.06	-0.15	0.43	0.08	0.01	-0.19
Obj9		0.34	0.37	0.19	0.21	0.15	0.52	-0.01	0.08	0.22	0.55	-0.12	0.04	0.19
Obj10		0.46	0.31	0.54	0.20	0.27	0.08	0.68	0.09	0.20	-0.05	0.69	0.02	0.13
Marg1		0.43	0.27	0.46	0.25	0.28	0.05	0.58	0.23	0.14	-0.07	0.58	0.18	0.06
Marg2		0.42	0.30	0.49	0.21	0.19	0.15	0.59	0.08	0.12	0.04	0.60	0.01	0.05
Marg3		0.43	0.32	0.46	0.19	0.22	0.20	0.56	-0.01	0.20	0.12	0.55	-0.08	0.15
Marg4		0.40	0.29	0.43	0.20	0.16	0.13	0.62	0.12	-0.06	0.02	0.64	0.05	-0.13
Marg5		0.48	0.37	0.45	0.27	0.24	0.27	0.48	0.06	0.27	0.19	0.45	-0.01	0.21
Marg6		0.32	0.24	0.39	0.08	0.18	0.09	0.61	-0.03	-0.03	-0.01	0.65	-0.09	-0.09
Marg7		0.35	0.32	0.23	0.28	0.14	0.25	0.15	0.55	-0.12	0.19	0.09	0.53	-0.21
Strong1		0.23	0.20	0.23	0.36	0.29	0.02	0.12	0.63	0.23	-0.05	0.05	0.63	0.15
Strong2		0.22	0.15	0.11	0.26	0.18	0.04	-0.03	0.57	0.12	0.00	-0.10	0.32	0.29
Strong3		0.26	0.13	0.24	0.17	0.31	0.01	0.14	0.01	0.70	-0.02	0.08	-0.02	0.70
Strong4		0.36	0.26	0.28	0.19	0.34	0.14	0.18	0.09	0.61	0.11	0.10	0.06	0.59
Strong5		0.32	0.20	0.23	0.29	0.31	-0.01	0.17	0.37	0.42	-0.07	0.11	0.37	0.37
Angry1		0.26	0.13	0.24	0.17	0.31	0.01	0.14	0.01	0.70	-0.02	0.08	-0.02	0.70
Angry2		0.36	0.26	0.28	0.19	0.34	0.14	0.18	0.09	0.61	0.11	0.10	0.06	0.59
Angry3		0.28	0.13	0.24	0.17	0.31	0.01	0.14	0.01	0.70	-0.02	0.08	-0.02	0.70

Figure 8.7: Image of a table of values from the item analysis and PCA solutions with orthogonal and oblique rotations

We are looking for:

- items that *load* higher on their own scales than they do on other scales
- when they are “close” or have a number of strong loadings, it means that it’s not going to discriminate well (think within-in scale discriminant validity).
  - if there are a number of these, there will likely be stronger correlations between subscales (indicating that the oblique rotation was an appropriate choice).
  - with low/no cross-loadings, this supports the choices of an orthogonal (uncorrelated) solution
- when the item has a strong, positive loading on its own scale, it supports within-scale convergent validity.

- similarities and differences across the item-analysis, PCA orthogonal, and PCA oblique solutions. Our biggest interest is in whether items change scale membership and/or have cross-loadings. This scale is performing extremely well with a great deal of stability
  - This, in part, is likely facilitated by the data simulation where we had the benefit of factors “telling” items where to load.

## 8.8 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The least complex is to change the random seed in the research and rework the problem demonstrated in the lesson. The most complex is to use data of your own. In either case, please plan to:

- Properly format and prepare the data.
- Conduct diagnostic tests to determine the suitability of the data for PCA.
- Conducting tests to guide the decisions about number of components to extract.
- Conducting orthogonal and oblique extractions (at least two each with different numbers of components).
- Selecting one solution and preparing an APA style results section (with table and figure).

### 8.8.1 Problem #1: Play around with this simulation.

Copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results. If PCA is new to you, perhaps you just change the number in “set.seed(240311)” from 240311 to something else. Your results should *parallel* those obtained in the lecture, making it easier for you to check your work as you go. Don’t be surprised if the factor loadings wiggle around a little. Do try to make sense of them.

### 8.8.2 Problem #2: Conduct a PCA with another simulated set of data in the OER.

The second option involves utilizing one of the simulated datasets available in this OER. Szymanski and Bissonette’s [2020] Perceptions of the LGBTQ College Campus Climate Scale: Development and psychometric evaluation and Keum et al.’s Gendered Racial Microaggressions Scale for Asian American Women [Keum et al., 2018] are ready for PCA analysis. The simulations are available in the chapters in which they are the featured vignette as well as in a simulations appendix at the end of the OER.

### 8.8.3 Problem #3: Try something entirely new.

Using data for which you have permission and access (e.g., IRB approved data you have collected or from your lab; data you simulate from a published article; data from an open science repository; data from other chapters in this OER), complete PCA. The data should allow for at least two (ideally three) components/subscales.

### 8.8.4 Grading Rubric

Using the lecture and workflow (chart) as a guide, please work through all the steps listed in the proposed assignment/grading rubric.

Assignment Component	Points Possible	Points Earned
1. Check and, if needed, format data	5	_____
2. Conduct and interpret the three diagnostic tests to determine if PCA is appropriate as an analysis (KMO, Bartlett's, determinant).	5	_____
3. Determine how many components to extract (e.g., scree plot, eigenvalues, theory).	5	_____
4. Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions.	5	_____
5. Conduct an oblique extraction and rotation with a minimum of two different factor extractions.	5	_____
6. Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest.	5	_____
7. APA style results section with table and figure of one of the solutions.	5	_____
8. Explanation to grader	5	_____
<b>Totals</b>	<b>40</b>	_____

## 8.9 Homeworked Example

### Screencast Link

For more information about the data used in this homeworked example, please refer to the description and codebook located at the end of the [introduction](#) in first volume of ReCentering Psych Stats.

As a brief review, this data is part of an IRB-approved study, with consent to use in teaching demonstrations and to be made available to the general public via the open science framework. Hence, it is appropriate to share in class. You will notice there are student- and teacher- IDs. These numbers are not connected to the SPU student ID. Rather, the subcontractor who does the course evals for SPU creates a third, not-SPU-related identifier.

This is the same dataset I have been using for many in-class demos. It's great for psychometrics because I actually created a three-factor solution from the institution's course evaluations. We'll get to walk through that process in this class.

Because this is an actual dataset, if you wish to work the problem along with me, you will need to download the data from **LINK TO DATASET**.

In this homewoRked example I will conduct a principal components analysis. My hope is that the results will support my solution of three dimensions: valued-by-the-student, traditional pedagogy, socially responsive pedagogy.

### 8.9.1 Check and, if needed, format data

```
big <- readRDS("ReC.rds")
```

With the next code I will create an item-level df with only the items used in the three scales.

```
library(tidyverse)
items <- big%>%
  dplyr::select (ValObjectives, IncrUnderstanding, IncrInterest, ClearResponsibilities, Effecti
```

Some of the analyses require non-missing data in the df.

```
items <- na.omit(items)
```

Let's check the structure of the data.

```
str(items)
```

```
Classes 'data.table' and 'data.frame': 267 obs. of 12 variables:
 $ ValObjectives      : int 5 5 4 4 5 5 5 4 5 3 ...
 $ IncrUnderstanding   : int 2 3 4 3 4 5 2 4 5 4 ...
 $ IncrInterest        : int 5 3 4 2 4 5 3 2 5 1 ...
 $ ClearResponsibilities: int 5 5 4 4 5 5 4 4 5 3 ...
 $ EffectiveAnswers    : int 5 3 5 3 5 4 3 2 3 3 ...
 $ Feedback             : int 5 3 4 2 5 5 4 4 5 2 ...
 $ ClearOrganization    : int 3 4 3 4 4 5 4 4 5 2 ...
 $ ClearPresentation    : int 4 4 4 2 5 4 4 4 5 2 ...
 $ MultPerspectives     : int 5 5 4 5 5 5 5 5 5 1 ...
 $ InclusvClassrm       : int 5 5 5 5 5 5 5 4 5 3 ...
 $ DEIintegration       : int 5 5 5 5 5 5 5 5 5 2 ...
 $ EquitableEval         : int 5 5 3 5 5 5 5 3 5 3 ...
 - attr(*, ".internal.selfref")=<externalptr>
 - attr(*, "na.action")= 'omit' Named int [1:43] 6 20 106 109 112 113 114 117 122 128 ...
 ..- attr(*, "names")= chr [1:43] "6" "20" "106" "109" ...
```

### 8.9.2 Conduct and interpret the three diagnostic tests to determine if PCA is appropriate as an analysis (KMO, Bartlett's, determinant)

#### 8.9.2.1 KMO

The Kaiser-Meyer-Olkin (KMO) index is an index of *sampling adequacy* to let us know if the sample size is sufficient relative to the statistical characteristics of the data.

General crieria (1974, Kaiser):

- bare minimum of .5

- values between .5 and .7 as mediocre
- values above .9 are superb

```
psych::KMO(items)
```

Kaiser-Meyer-Olkin factor adequacy  
 Call: psych::KMO(r = items)  
 Overall MSA = 0.91  
 MSA for each item =

	ValObjectives	IncrUnderstanding	IncrInterest
	0.94	0.89	0.89
ClearResponsibilities		EffectiveAnswers	Feedback
	0.91	0.93	0.94
ClearOrganization		ClearPresentation	MultPerspectives
	0.94	0.91	0.93
InclusvClassrm		DEIntegration	EquitableEval
	0.86	0.78	0.95

With a KMO of 0.91, the data seems appropriate to continue with the PCA.

### 8.9.2.2 Bartlett's

Barlett's test let's us know if the matrix is an *identity matrix* (i.e., where elements on the off-diagonal would be 0.0 and elements on the diagonal would be 1.0). Stated another way – items only correlate with “themselves” and not other variables.

When  $p < 0.05$  the matrix is not an identity matrix. That is, there are some relationships between variables that can be analyzed.

```
psych::cortest.bartlett(items)
```

R was not square, finding R from data

```
$chisq
[1] 1897.769

$p.value
[1] 0

$df
[1] 66
```

The Barlett's test,  $\chi^2(66) = 1897.77, p < 0.001$ , indicating that the correlation matrix is not an identity matrix and, on that dimension, is suitable for analysis.

### 8.9.2.3 Determinant

Multicollinearity or singularity is diagnosed by the determinant. The determinant should be greater than 0.00001. If smaller, then there may be an issue with multicollinearity (variables that are too highly correlated) or singularity (variables that are perfectly correlated).

```
items <- na.omit(items)
det(cor(items))
```

```
[1] 0.0006985496
```

The value of the determinant is 0.0007; greater than 0.00001. We are not concerned with multicollinearity or singularity.

Summary from data screening:

Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was 0.91, indicating acceptable sampling adequacy. The Bartlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(66) = 1897.77, p < 0.001$  indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.0007 and, again, indicated that our data was suitable for the analysis.

### 8.9.3 Determine how many components to extract (e.g., scree plot, eigenvalues, theory)

Step #1: creating a principal components model with the same number of components as items

```
pca1 <- psych::principal(items, nfactors=length(items), rotate="none")# using raw data and let
pca1
```

#### Principal Components Analysis

```
Call: psych::principal(r = items, nfactors = length(items), rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
          PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9
ValObjectives   0.57 -0.13  0.42  0.68 -0.08  0.03  0.03 -0.06 -0.04
IncrUnderstanding  0.68 -0.37  0.39 -0.28  0.08  0.09 -0.12  0.25 -0.04
IncrInterest     0.73 -0.19  0.41 -0.17  0.32 -0.05  0.06 -0.12  0.03
```

ClearResponsibilities	0.81	-0.08	-0.37	0.04	-0.20	0.02	-0.03	0.09	-0.10
EffectiveAnswers	0.80	-0.16	-0.20	-0.09	-0.03	0.05	0.48	-0.05	0.07
Feedback	0.77	0.06	-0.28	0.11	0.32	-0.29	0.01	0.09	-0.31
ClearOrganization	0.79	-0.27	-0.12	0.04	-0.22	-0.22	-0.20	0.19	0.20
ClearPresentation	0.85	-0.21	0.00	-0.11	-0.24	0.01	0.05	-0.18	0.04
MultPerspectives	0.79	0.26	-0.11	-0.05	0.13	-0.14	-0.23	-0.35	0.17
InclusvClassrm	0.66	0.50	0.24	-0.20	-0.30	0.11	-0.06	-0.07	-0.28
DEIIntegration	0.50	0.75	0.19	0.03	0.02	-0.12	0.14	0.25	0.21
EquitableEval	0.73	0.13	-0.27	0.13	0.22	0.52	-0.12	0.07	0.07
	PC10	PC11	PC12	h2				u2	com
ValObjectives	0.05	0.02	0.04	1	-0.00000000000000022	2.9			
IncrUnderstanding	0.27	0.00	0.07	1	0.000000000000000389	3.7			
IncrInterest	-0.28	0.15	-0.09	1	0.000000000000000111	3.0			
ClearResponsibilities	0.05	0.36	-0.11	1	0.00000000000000089	2.2			
EffectiveAnswers	0.03	-0.01	0.19	1	0.00000000000000078	2.1			
Feedback	0.02	-0.14	-0.02	1	0.00000000000000089	2.6			
ClearOrganization	-0.23	-0.09	0.11	1	0.00000000000000033	2.5			
ClearPresentation	0.08	-0.21	-0.28	1	0.000000000000000122	1.9			
MultPerspectives	0.18	0.06	0.12	1	0.000000000000000100	2.4			
InclusvClassrm	-0.13	-0.03	0.12	1	0.000000000000000056	3.7			
DEIIntegration	0.06	0.01	-0.11	1	0.000000000000000033	2.6			
EquitableEval	-0.08	-0.09	-0.02	1	0.000000000000000011	2.7			

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
SS loadings	6.38	1.23	0.95	0.67	0.52	0.47	0.39	0.37	0.32	0.27	0.24
Proportion Var	0.53	0.10	0.08	0.06	0.04	0.04	0.03	0.03	0.03	0.02	0.02
Cumulative Var	0.53	0.63	0.71	0.77	0.81	0.85	0.88	0.91	0.94	0.96	0.98
Proportion Explained	0.53	0.10	0.08	0.06	0.04	0.04	0.03	0.03	0.03	0.02	0.02
Cumulative Proportion	0.53	0.63	0.71	0.77	0.81	0.85	0.88	0.91	0.94	0.96	0.98
	PC12										
SS loadings	0.20										
Proportion Var	0.02										
Cumulative Var	1.00										
Proportion Explained	0.02										
Cumulative Proportion	1.00										

Mean item complexity = 2.7

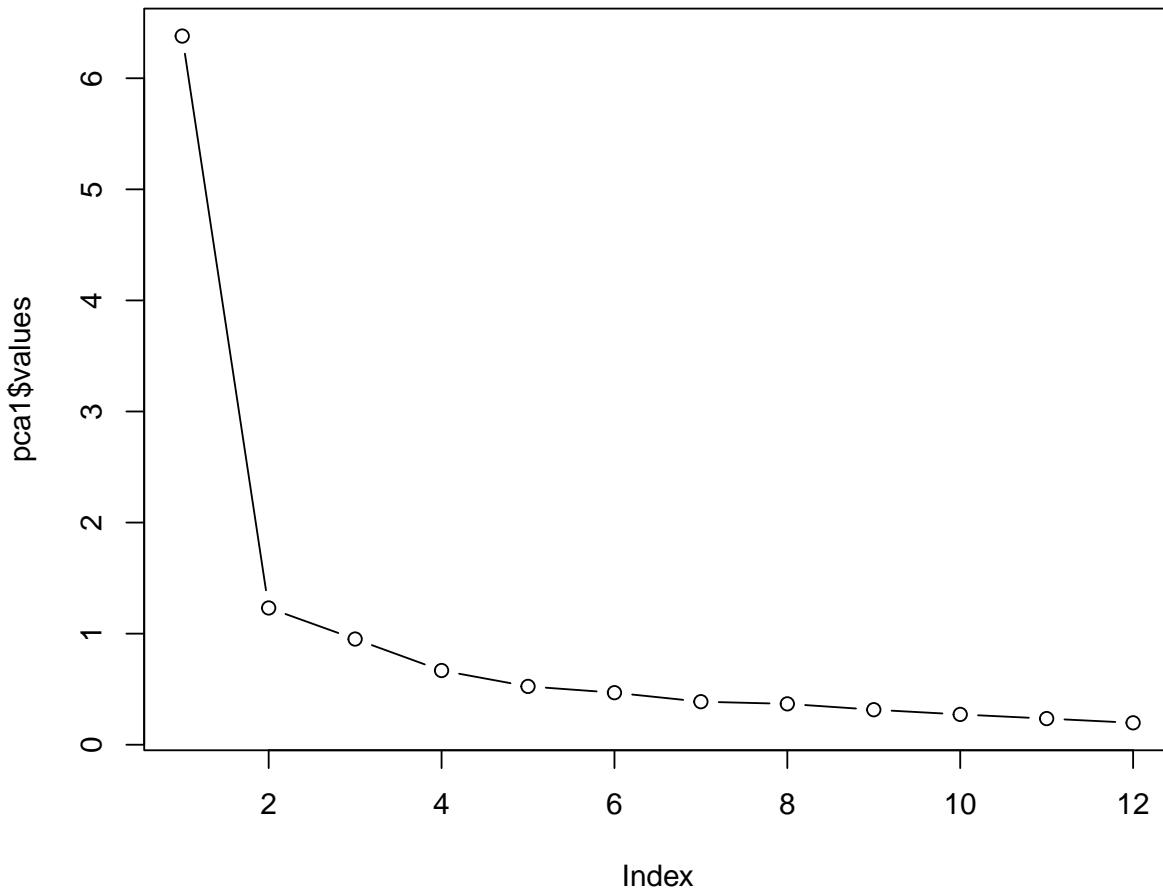
Test of the hypothesis that 12 components are sufficient.

The root mean square of the residuals (RMSR) is 0  
with the empirical chi square 0 with prob < NA

Fit based upon off diagonal values = 1

The eigenvalue-greater-than-one criteria suggests 2 factors (but the third component has an SS-loading of .95 – it's close to three).

```
plot(pca1$values, type = "b")
```



The scree plot looks like one factor.

Ugh.

- I want 3 factors (we could think of this as a priori theory); would account for 71% of variance.
- Eigenvalues-greater-than-one criteria suggests two; could account for 63% of variance.
- Scree plot suggests 1 (would account for 53% of variance)

*Note:* The lecture has more on evaluating communalities and uniquenesses and how this information can also inform the number of components we want to extract. Because it is easy to get lost (very lost) I will skip over this for now. If you were to create a measure and use PCA as an exploratory approach to understanding the dimensionality of an instrument, you would likely want to investigate further and report on these.

### 8.9.4 Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions

#### An orthogonal two factor solution

```
pcaORTH2f <- psych::principal(items, nfactors = 2, rotate = "varimax")
pcaORTH2f
```

Principal Components Analysis  
 Call: psych::principal(r = items, nfactors = 2, rotate = "varimax")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	RC1	RC2	h2	u2	com
ValObjectives	0.55	0.19	0.34	0.66	1.2
IncrUnderstanding	0.77	0.05	0.59	0.41	1.0
IncrInterest	0.72	0.23	0.57	0.43	1.2
ClearResponsibilities	0.73	0.36	0.66	0.34	1.5
EffectiveAnswers	0.76	0.29	0.67	0.33	1.3
Feedback	0.62	0.46	0.59	0.41	1.8
ClearOrganization	0.81	0.18	0.69	0.31	1.1
ClearPresentation	0.83	0.27	0.76	0.24	1.2
MultPerspectives	0.53	0.64	0.70	0.30	1.9
InclusivClassrm	0.29	0.77	0.68	0.32	1.3
DEIIntegration	0.03	0.90	0.80	0.20	1.0
EquitableEval	0.55	0.50	0.55	0.45	2.0

	RC1	RC2
SS loadings	4.93	2.68
Proportion Var	0.41	0.22
Cumulative Var	0.41	0.63
Proportion Explained	0.65	0.35
Cumulative Proportion	0.65	1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07  
 with the empirical chi square 170.34 with prob < 0.0000000000000045

Fit based upon off diagonal values = 0.98

Sorting the scores into a table can help see the results more clearly. The “cut = #” command will not show the factor scores for factor loading < .30. I would do this “to see”, but I would include all the values in an APA style table.

```
pca_tableORT2f <- psych::print.psych(pcaORTH2f, cut = 0.3, sort=TRUE)
```

Principal Components Analysis

Call: psych::principal(r = items, nfactors = 2, rotate = "varimax")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	item	RC1	RC2	h2	u2	com
ClearPresentation	8	0.83		0.76	0.24	1.2
ClearOrganization	7	0.81		0.69	0.31	1.1
IncrUnderstanding	2	0.77		0.59	0.41	1.0
EffectiveAnswers	5	0.76		0.67	0.33	1.3
ClearResponsibilities	4	0.73	0.36	0.66	0.34	1.5
IncrInterest	3	0.72		0.57	0.43	1.2
Feedback	6	0.62	0.46	0.59	0.41	1.8
ValObjectives	1	0.55		0.34	0.66	1.2
EquitableEval	12	0.55	0.50	0.55	0.45	2.0
DEIIntegration	11		0.90	0.80	0.20	1.0
InclusvClassrm	10		0.77	0.68	0.32	1.3
MultPerspectives	9	0.53	0.64	0.70	0.30	1.9

	RC1	RC2
SS loadings	4.93	2.68
Proportion Var	0.41	0.22
Cumulative Var	0.41	0.63
Proportion Explained	0.65	0.35
Cumulative Proportion	0.65	1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 components are sufficient.

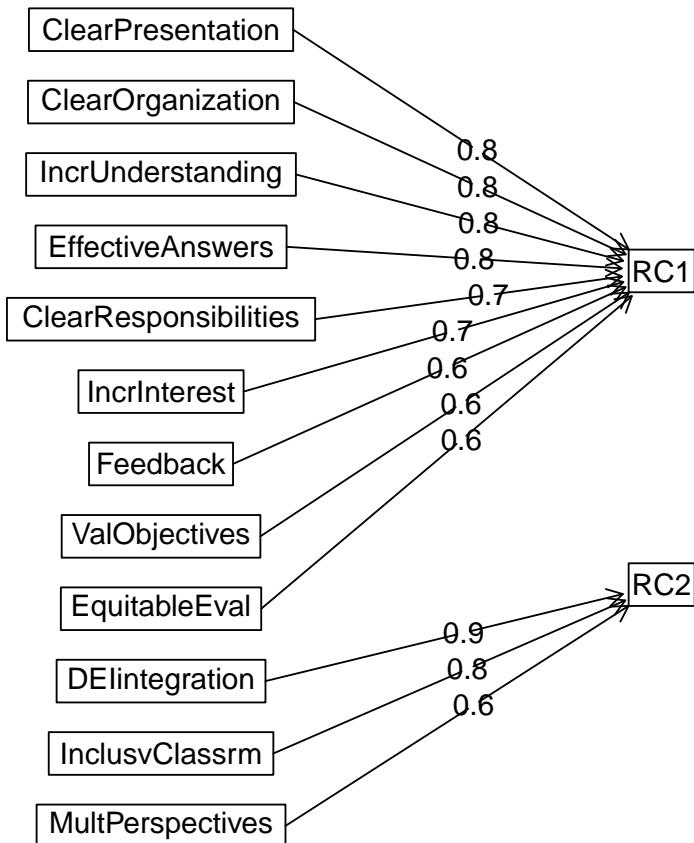
The root mean square of the residuals (RMSR) is 0.07  
 with the empirical chi square 170.34 with prob < 0.0000000000000045

Fit based upon off diagonal values = 0.98

F1: Includes everything else. F2: Includes the SCR items (although MultPerspectives cross-loads onto F1; Similarly, EquitableEval is on F1)

```
psych::fa.diagram(pcaORTH2f)
```

## Components Analysis



Plotting these figures from the program can facilitate conceptual understanding of what is going on – and can be a “check” to your work.

In the lecture I made a “biggish deal” about PCA being *components* (not *factor*) analysis. Although the two approaches can lead to similar results/conclusions, there are some significant differences “under the hood.” PCA can be thought of more as regression where the items predict the component. Consequently, the arrows go *from* the item, *to* the component. Starting with the next lesson, the arrows will go from the factor to the item – because the factors (or latent variables) are assumed to predict the scores on the items (i.e., “depression” would predict how someone rates items that assess hopelessness, sleep, anhedonia, and so forth).

### An orthogonal three factor solution

```
pcaORTH3f <- psych::principal(items, nfactors = 3, rotate = "varimax")
pcaORTH3f
```

```
Principal Components Analysis
Call: psych::principal(r = items, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	RC1	RC3	RC2	h2	u2	com
ValObjectives	0.16	0.67	0.21	0.52	0.48	1.3
IncrUnderstanding	0.29	0.81	0.04	0.75	0.25	1.3
IncrInterest	0.30	0.78	0.23	0.74	0.26	1.5
ClearResponsibilities	0.85	0.22	0.14	0.80	0.20	1.2
EffectiveAnswers	0.75	0.37	0.12	0.71	0.29	1.5
Feedback	0.75	0.19	0.28	0.67	0.33	1.4
ClearOrganization	0.70	0.47	0.03	0.71	0.29	1.8
ClearPresentation	0.65	0.56	0.14	0.76	0.24	2.1
MultPerspectives	0.63	0.24	0.51	0.71	0.29	2.2
InclusvClassrm	0.26	0.30	0.76	0.74	0.26	1.6
DEIIntegration	0.15	0.07	0.90	0.84	0.16	1.1
EquitableEval	0.70	0.15	0.33	0.62	0.38	1.5

	RC1	RC3	RC2
SS loadings	3.93	2.64	1.99
Proportion Var	0.33	0.22	0.17
Cumulative Var	0.33	0.55	0.71
Proportion Explained	0.46	0.31	0.23
Cumulative Proportion	0.46	0.77	1.00

Mean item complexity = 1.5

Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.06  
with the empirical chi square 115.69 with prob < 0.000000000041

Fit based upon off diagonal values = 0.99

```
pca_tableORT3f <- psych::print.psych(pcaORTH3f, cut = 0.3, sort=TRUE)
```

```
Principal Components Analysis
Call: psych::principal(r = items, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
          item  RC1   RC3   RC2   h2   u2 com
ClearResponsibilities  4  0.85           0.80 0.20 1.2
EffectiveAnswers       5  0.75  0.37           0.71 0.29 1.5
Feedback                6  0.75           0.67 0.33 1.4
EquitableEval          12 0.70           0.33 0.62 0.38 1.5
ClearOrganization       7  0.70  0.47           0.71 0.29 1.8
ClearPresentation        8  0.65  0.56           0.76 0.24 2.1
MultPerspectives        9  0.63           0.51 0.71 0.29 2.2
IncrUnderstanding        2    0.81           0.75 0.25 1.3
IncrInterest             3    0.78           0.74 0.26 1.5
ValObjectives            1    0.67           0.52 0.48 1.3
DEIIntegration           11           0.90 0.84 0.16 1.1
InclusvClassrm          10           0.30 0.76 0.74 0.26 1.6
```

	RC1	RC3	RC2
SS loadings	3.93	2.64	1.99
Proportion Var	0.33	0.22	0.17
Cumulative Var	0.33	0.55	0.71
Proportion Explained	0.46	0.31	0.23
Cumulative Proportion	0.46	0.77	1.00

Mean item complexity = 1.5

Test of the hypothesis that 3 components are sufficient.

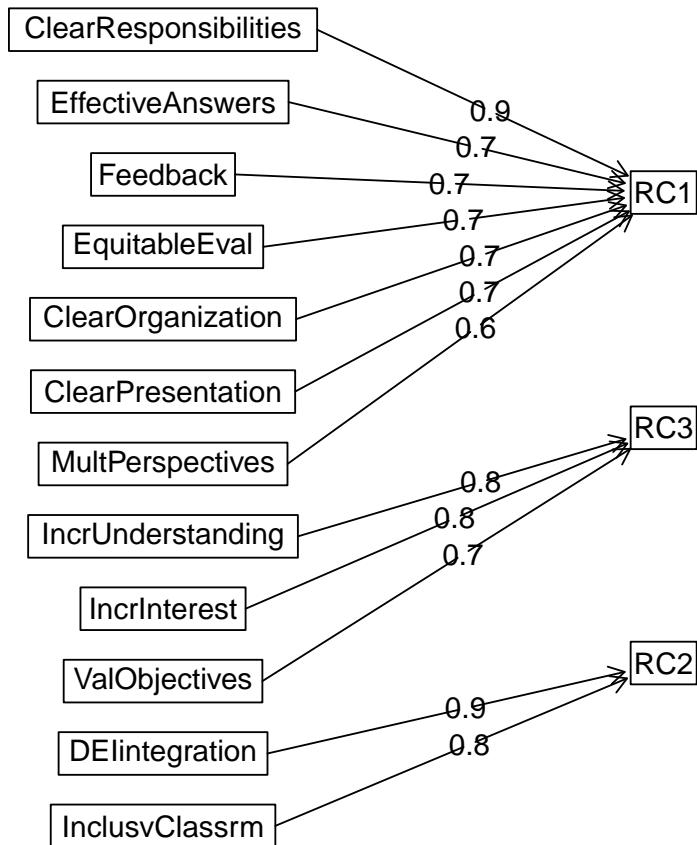
The root mean square of the residuals (RMSR) is 0.06  
with the empirical chi square 115.69 with prob < 0.000000000041

Fit based upon off diagonal values = 0.99

F1: Traditional Pedagogy F2: Valued-by-Me F3: SCRPed—except Equitable Eval \* MultPerspectives are on TradPed; MultPerspectives cross-load

```
psych::fa.diagram(pcaORTH3f)
```

## Components Analysis



The three factor solution gets really close to my goals of (a) traditional pedagogy, (b) valued by the student, and (c) socially responsive pedagogy. The trouble is that I would prefer “multiple perspectives” to load with the socially responsive pedagogy factor.

### 8.9.5 Conduct an oblique extraction and rotation with a minimum of two different factor extractions

#### An oblique two factor solution

```
pca0BL2f <- psych::principal(items, nfactors = 2, rotate = "oblimin")
pca0BL2f
```

```
Principal Components Analysis
Call: psych::principal(r = items, nfactors = 2, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
          TC1   TC2   h2   u2 com
ValObjectives    0.58  0.01  0.34  0.66  1.0
```

IncrUnderstanding	0.84	-0.21	0.59	0.41	1.1
IncrInterest	0.76	0.00	0.57	0.43	1.0
ClearResponsibilities	0.75	0.13	0.66	0.34	1.1
EffectiveAnswers	0.80	0.05	0.67	0.33	1.0
Feedback	0.61	0.27	0.59	0.41	1.4
ClearOrganization	0.86	-0.08	0.69	0.31	1.0
ClearPresentation	0.87	0.00	0.76	0.24	1.0
MultPerspectives	0.50	0.49	0.70	0.30	2.0
InclusvClassrm	0.21	0.71	0.68	0.32	1.2
DEIIntegration	-0.10	0.93	0.80	0.20	1.0
EquitableEval	0.53	0.34	0.55	0.45	1.7

	TC1	TC2
SS loadings	5.50	2.11
Proportion Var	0.46	0.18
Cumulative Var	0.46	0.63
Proportion Explained	0.72	0.28
Cumulative Proportion	0.72	1.00

With component correlations of

TC1	TC2
TC1	1.00 0.43
TC2	0.43 1.00

Mean item complexity = 1.2

Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07  
with the empirical chi square 170.34 with prob < 0.0000000000000045

Fit based upon off diagonal values = 0.98

```
pca_table0BL2f <- psych::print.psych(pca0BL2f, cut = 0.3, sort=TRUE)
```

#### Principal Components Analysis

Call: psych::principal(r = items, nfactors = 2, rotate = "oblimin")  
Standardized loadings (pattern matrix) based upon correlation matrix

	item	TC1	TC2	h2	u2	com
ClearPresentation	8	0.87		0.76	0.24	1.0
ClearOrganization	7	0.86		0.69	0.31	1.0
IncrUnderstanding	2	0.84		0.59	0.41	1.1
EffectiveAnswers	5	0.80		0.67	0.33	1.0
IncrInterest	3	0.76		0.57	0.43	1.0
ClearResponsibilities	4	0.75		0.66	0.34	1.1
Feedback	6	0.61		0.59	0.41	1.4
ValObjectives	1	0.58		0.34	0.66	1.0
EquitableEval	12	0.53	0.34	0.55	0.45	1.7

MultPerspectives	9	0.50	0.49	0.70	0.30	2.0
DEIintegration	11		0.93	0.80	0.20	1.0
InclusvClassrm	10		0.71	0.68	0.32	1.2

	TC1	TC2
SS loadings	5.50	2.11
Proportion Var	0.46	0.18
Cumulative Var	0.46	0.63
Proportion Explained	0.72	0.28
Cumulative Proportion	0.72	1.00

With component correlations of

TC1	TC2
TC1	1.00 0.43
TC2	0.43 1.00

Mean item complexity = 1.2

Test of the hypothesis that 2 components are sufficient.

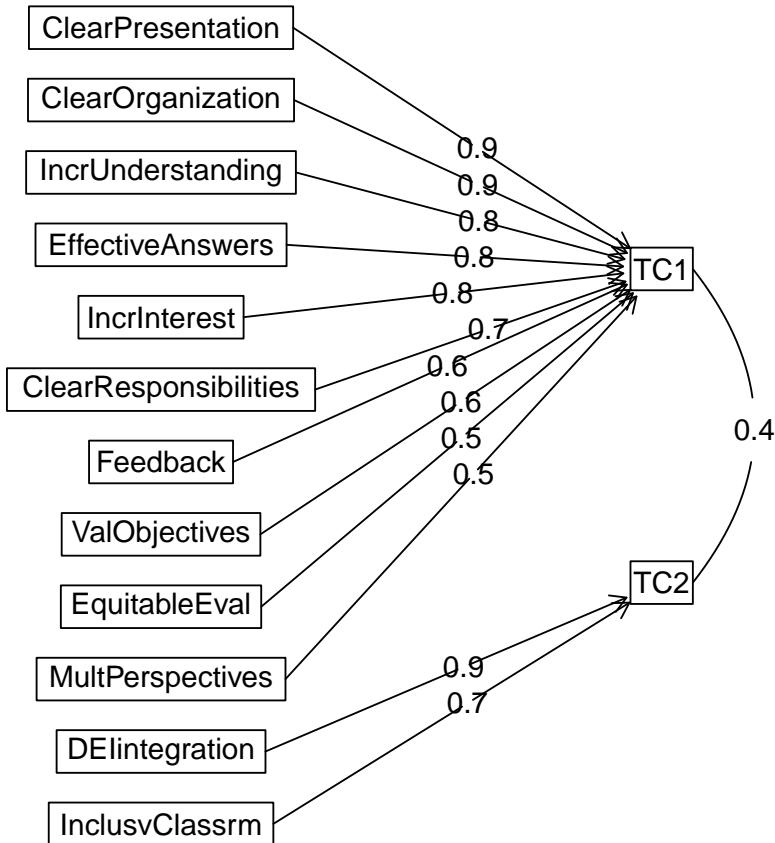
The root mean square of the residuals (RMSR) is 0.07  
 with the empirical chi square 170.34 with prob < 0.00000000000000045

Fit based upon off diagonal values = 0.98

Fairly similar results to the orthogonal variation of this – with EquitableEval and MultPerspectives cross-loading, with stronger loadings on the TradPed/Valued dimension.

```
psych::fa.diagram(pca0BL2f)
```

## Components Analysis



The curved curved line and value between TC1 and TC2 illustrates that in the oblique solution the components are allowed to correlate. There was no such path on the orthogonal figures. This is because the rotation required the components to be uncorrelated.

### An oblique three factor solution

```
pca0BL3f <- psych::principal(items, nfactors = 3, rotate = "oblimin")
pca0BL3f
```

```
Principal Components Analysis
Call: psych::principal(r = items, nfactors = 3, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
          TC1    TC3    TC2    h2   u2 com
ValObjectives -0.08  0.71  0.15  0.52  0.48 1.1
IncrUnderstanding  0.06  0.84 -0.07  0.75  0.25 1.0
IncrInterest      0.05  0.79  0.13  0.74  0.26 1.1
ClearResponsibilities  0.95 -0.06 -0.05  0.80  0.20 1.0
EffectiveAnswers     0.76  0.15 -0.06  0.71  0.29 1.1
```

Feedback	0.80	-0.06	0.12	0.67	0.33	1.1
ClearOrganization	0.68	0.29	-0.15	0.71	0.29	1.5
ClearPresentation	0.57	0.41	-0.02	0.76	0.24	1.8
MultPerspectives	0.59	0.03	0.39	0.71	0.29	1.7
InclusvClassrm	0.08	0.22	0.73	0.74	0.26	1.2
DEIIntegration	0.00	-0.02	0.92	0.84	0.16	1.0
EquitableEval	0.75	-0.10	0.18	0.62	0.38	1.2

	TC1	TC3	TC2
SS loadings	4.23	2.50	1.83
Proportion Var	0.35	0.21	0.15
Cumulative Var	0.35	0.56	0.71
Proportion Explained	0.49	0.29	0.21
Cumulative Proportion	0.49	0.79	1.00

With component correlations of

	TC1	TC3	TC2
TC1	1.00	0.58	0.39
TC3	0.58	1.00	0.25
TC2	0.39	0.25	1.00

Mean item complexity = 1.2

Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.06  
with the empirical chi square 115.69 with prob < 0.000000000041

Fit based upon off diagonal values = 0.99

```
pca_table0BL3f <- psych::print.psych(pca0BL3f, cut = 0.3, sort=TRUE)
```

#### Principal Components Analysis

Call: psych::principal(r = items, nfactors = 3, rotate = "oblimin")  
Standardized loadings (pattern matrix) based upon correlation matrix

	item	TC1	TC3	TC2	h2	u2	com
ClearResponsibilities	4	0.95			0.80	0.20	1.0
Feedback	6	0.80			0.67	0.33	1.1
EffectiveAnswers	5	0.76			0.71	0.29	1.1
EquitableEval	12	0.75			0.62	0.38	1.2
ClearOrganization	7	0.68			0.71	0.29	1.5
MultPerspectives	9	0.59		0.39	0.71	0.29	1.7
ClearPresentation	8	0.57	0.41		0.76	0.24	1.8
IncrUnderstanding	2		0.84		0.75	0.25	1.0
IncrInterest	3		0.79		0.74	0.26	1.1
ValObjectives	1		0.71		0.52	0.48	1.1
DEIIntegration	11			0.92	0.84	0.16	1.0
InclusvClassrm	10			0.73	0.74	0.26	1.2

	TC1	TC3	TC2
SS loadings	4.23	2.50	1.83
Proportion Var	0.35	0.21	0.15
Cumulative Var	0.35	0.56	0.71
Proportion Explained	0.49	0.29	0.21
Cumulative Proportion	0.49	0.79	1.00

With component correlations of

	TC1	TC3	TC2
TC1	1.00	0.58	0.39
TC3	0.58	1.00	0.25
TC2	0.39	0.25	1.00

Mean item complexity = 1.2

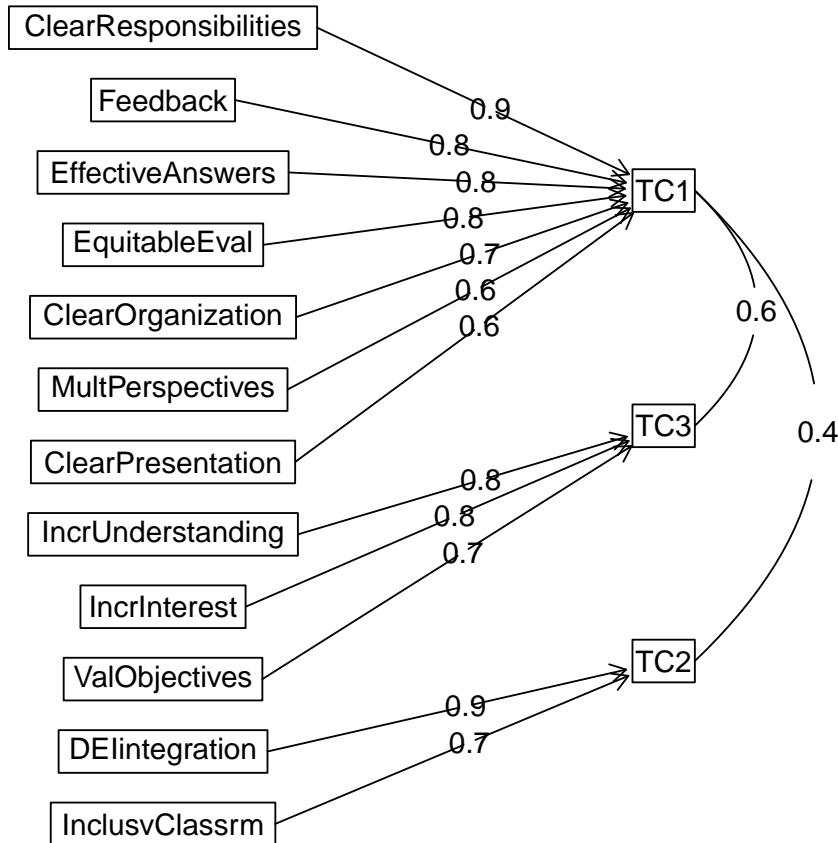
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.06  
with the empirical chi square 115.69 with prob < 0.000000000041

Fit based upon off diagonal values = 0.99

```
psych::fa.diagram(pca0BL3f)
```

## Components Analysis



The results are quite similar to the orthogonal solution.

### 8.9.6 Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest

From the oblique output we see that the correlations between the three subscales range from 0.25 to 0.58. These are high. Therefore, I will choose a 3-component, oblique, solution.

### 8.9.7 APA style results section with table and figure of one of the solutions

The dimensionality of the 12 course evaluation items was analyzed using principal components analysis. First, data were screened to determine the suitability of the data for this analyses. Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that

component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was 0.91, indicating acceptable sampling adequacy. The Barlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(66) = 1897.77, p < 0.001$  indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.0007 and, again, indicated that our data was suitable for the analysis.

Four criteria were used to determine the number of components to extract: a priori theory, the scree test, the eigenvalue-greater-than-one criteria, and the interpretability of the solution. Kaiser's eigenvalue-greater-than-one criteria suggested two components, and, in combination explained 63% of the variance. The inflexion in the scree plot justified retaining one component. A priorily, we researchers were expecting three components – which would explain 71% of the variance. Correspondingly, we investigated two and three component solutions with orthogonal (varimax) and oblique (oblimin) procedures. Given the significant correlations (ranging from .25 to .58) and the correspondence of items loading on the a priorili hypothesized components, we determined that an oblique, three-component, solution was most appropriate.

The rotated solution, as shown in Table 1 and Figure 1, yielded three interpretable components, each listed with the proportion of variance accounted for: traditional pedagogy (35%), valued-by-me (21%), and socially and culturally responsive pedagogy (15%).

Regarding the Table 1, I would include a table with ALL the values, bolding those with component membership. This is easy, though, because we can export it to a .csv file and

```
pcaOBL3fb <- psych::principal(items, nfactors = 3, rotate = "oblimin")
pca_tableOBL3fb <- psych::print.psych(pcaOBL3fb, sort=TRUE)
```

	item	TC1	TC3	TC2	h2	u2	com
ClearResponsibilities	4	0.95	-0.06	-0.05	0.80	0.20	1.0
Feedback	6	0.80	-0.06	0.12	0.67	0.33	1.1
EffectiveAnswers	5	0.76	0.15	-0.06	0.71	0.29	1.1
EquitableEval	12	0.75	-0.10	0.18	0.62	0.38	1.2
ClearOrganization	7	0.68	0.29	-0.15	0.71	0.29	1.5
MultPerspectives	9	0.59	0.03	0.39	0.71	0.29	1.7
ClearPresentation	8	0.57	0.41	-0.02	0.76	0.24	1.8
IncrUnderstanding	2	0.06	0.84	-0.07	0.75	0.25	1.0
IncrInterest	3	0.05	0.79	0.13	0.74	0.26	1.1
ValObjectives	1	-0.08	0.71	0.15	0.52	0.48	1.1

DEIintegration	11	0.00	-0.02	0.92	0.84	0.16	1.0
InclusvClassrm	10	0.08	0.22	0.73	0.74	0.26	1.2

	TC1	TC3	TC2
SS loadings	4.23	2.50	1.83
Proportion Var	0.35	0.21	0.15
Cumulative Var	0.35	0.56	0.71
Proportion Explained	0.49	0.29	0.21
Cumulative Proportion	0.49	0.79	1.00

With component correlations of

TC1	TC3	TC2	
TC1	1.00	0.58	0.39
TC3	0.58	1.00	0.25
TC2	0.39	0.25	1.00

Mean item complexity = 1.2

Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.06  
with the empirical chi square 115.69 with prob < 0.000000000041

Fit based upon off diagonal values = 0.99

```
pcaOBL3fb_table <- round(pcaOBL3fb$loadings,3)
write.table(pcaOBL3fb_table, file="pcaOBL3f_table.csv", sep=",", col.names=TRUE, row.names=TRUE)
pcaOBL3fb_table
```

Loadings:

	TC1	TC3	TC2
ValObjectives		0.712	0.151
IncrUnderstanding		0.844	
IncrInterest		0.787	0.132
ClearResponsibilities	0.947		
EffectiveAnswers	0.764	0.154	
Feedback	0.800		0.119
ClearOrganization	0.685	0.293	-0.149
ClearPresentation	0.574	0.413	
MultPerspectives	0.593		0.391
InclusvClassrm		0.218	0.730
DEIintegration			0.921
EquitableEval	0.751		0.184

	TC1	TC3	TC2
SS loadings	3.854	2.185	1.655
Proportion Var	0.321	0.182	0.138
Cumulative Var	0.321	0.503	0.641

**8.9.8 Explanation to grader**

# Chapter 9

# Principal Axis Factoring

## [Screencasted Lecture Link](#)

This is the second week of *exploratory* principal components analysis (PCA) and factor analysis (EFA). This time the focus is on actual *factor analysis*. There are numerous approaches. I will be demonstrating principal axis factoring (PAF).

## 9.1 Navigating this Lesson

There is about an hour-and-a-half of lecture. If you work through the materials with me it would be plan for an additional two hours.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

### 9.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Distinguish between PCA and EFA on several levels:
  - recognize PCA and EFA from a path diagram
  - define keywords associated with each: factor loadings, linear components, describe v. explain.
- Recognize/define an identity matrix – what test would you use to diagnose it?
- Recognize/define multicollinearity and singularity – what test would you use to diagnose it?
- Describe the desired pattern of “loadings” (i.e., the relative weights of an item on its own scale compared to other scales)
- Compare the results from item analysis, PCA, PAF, and omega.

### 9.1.2 Planning for Practice

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. Whichever you choose, it would be terrific if you used the same dataframe across as many psychometrics lessons as possible so you can compare the results.

The least complex is to change the random seed and rework the problem demonstrated in the lesson. The results *should* map onto the ones obtained in the lecture.

The second option involves utilizing one of the simulated datasets available in this OER. Szymanski and Bissonette's [2020] Perceptions of the LGBTQ College Campus Climate Scale: Development and Psychometric Evaluation was used as the research vignette for the validity, reliability, and item analysis lessons. Although I switched vignettes, the Szymanski and Bissonette example is ready for PCA.

As a third option, you are welcome to use data to which you have access and is suitable for PCA. These could include other vignettes from this OER, other simulated data, or your own data (presuming you have permission to use it). In either case, please plan to:

- Properly format and prepare the data.
- Conduct diagnostic tests to determine the suitability of the data for PCA.
- Conducting tests to guide the decisions about number of components to extract.
- Conducting orthogonal and oblique extractions (at least two each with different numbers of components).
- Selecting one solution and preparing an APA style results section (with table and figure).
- Compare your results in light of any other psychometrics lessons where you have used this data (especially the **item analysis** and **PCA** lessons).

### 9.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

- Revelle, William. (n.d.). Chapter 6: Constructs, components, and factor models. In *An introduction to psychometric theory with applications in R*. Retrieved from <https://personality-project.org/r/book/#chapter6>
  - pp. 150 to 167. Stop at “Non-Simple Structure Solutions: The Simplex and Circumplex.”
  - A simultaneously theoretical review of psychometric theory while working with R and data to understand the concepts.
- Revelle, W. (2019). *How To: Use the psych package for Factor Analysis and data reduction*.
  - Treat as reference. Pages 13 through 24 provide technical information about what we are doing.

### 9.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
# will install the package if not already installed
# if(!require(psych)){install.packages('psych')}
# if(!require(tidyverse)){install.packages('tidyverse')}
# if(!require(MASS)){install.packages('MASS')}
# if(!require(sjstats)){install.packages('sjstats')}
# if(!require(apaTables)){install.packages('apaTables')}
# if(!require(qualtrics)){install.packages('qualtrics')}
```

## 9.2 Exploratory Factor Analysis (with a quick contrast to PCA)

Whereas principal components analysis (PCA) is a regression analysis technique, principal factor analysis is “...a latent variable model” [?].

Exploratory factor analysis has a rich history. In 1904, Spearman used it for a single factor. In 1947, Thurstone generalized it to multiple factors. Factor analysis is frequently used and controversial.

Factor analysis and principal components are commonly confused:

### Principal components

- linear sums of variables,
- solved with an eigenvalue or singular decomposition
- represents a  $n \times n$  matrix in terms of the first  $k$  components and attempts to reproduce all of the  $R$  matrix.
- paths point from the items to a total scale score – all represented as observed/manifest (square) variables

### Factor analysis

- linear sums of unknown factors
- estimated as best fitting solutions, normally through iterative procedures.
- Controversial because
  - at the *structural* level (i.e., covariance or correlation matrix), there are normally more observed variables than parameters to estimate them and the procedure seeks to find the best fitting solution using ordinary least squares, weighted least squares, or maximum likelihood
  - at the *data* level, the model is indeterminate, although scores can be estimated
  - this leads some to argue for using principal components; but fans of factor analysis suggest that it is useful for theory construction and evaluation
- attempts to model only the *common* part of the matrix, which means all of the off-diagonal elements and the common part of the diagonal (the *communalities*); the *uniquenesses* are the non-common (leftover) part
- Stated another way, the factor model partitions the correlation or covariance matrix into
  - *common factors*,  $FF'$ , and
  - that which is *unique*,  $U^2$  (the diagonal matrix of *uniquenesses*)

- paths point from the latent variable (LV) representing the factor (oval) to the items (squares) illustrating that the factor/LV “causes” the item’s score

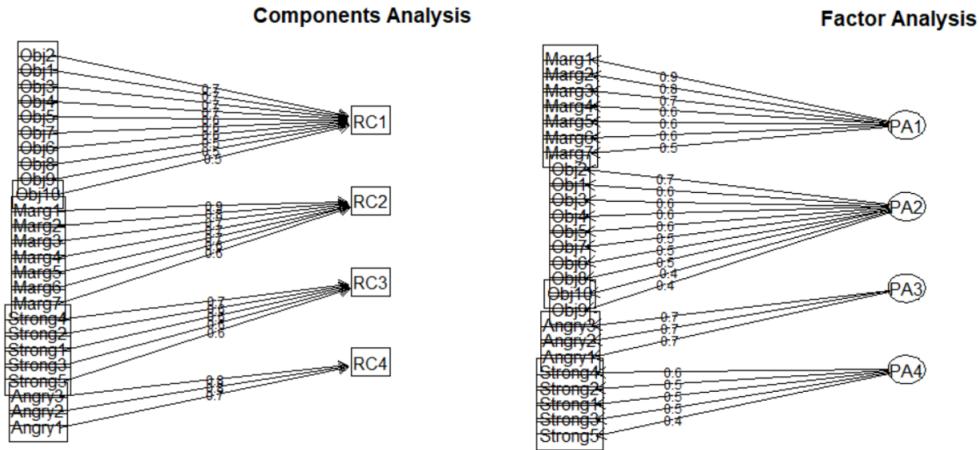


Figure 9.1: Comparison of path models for PCA and EFA

Our focus today is on the PAF approach to scale construction. By utilizing the same research vignette as in the [PCA lesson](#), we can identify similarities in differences in the approach, results, and interpretation. Let’s first take a look at the workflow for PAF.

### 9.3 PAF Workflow

Below is a screenshot of the workflow. The original document is located in the [Github site](#) that hosts the ReCentering Psych Stats: Psychometrics OER. You may find it refreshing that, with the exception of the change from “components” to “factors,” the workflow for PCA and PAF are quite similar.

Steps in the process include:

- Creating an items only dataframe where all items are scaled in the same direction (i.e., negatively worded items are reverse-scored).
- Conducting tests that assess the statistical assumptions of PAF to ensure that the data is appropriate for PAF.
- Determining the number of factors (think “subscale”) to extract.
- Conducting the factor extraction – this process will likely occur iteratively,
  - exploring orthogonal (uncorrelated/independent) and oblique (correlated) factors, and
  - changing the number of factors to extract

Because the intended audience for the ReCentering Psych Stats OER is the scientist-practitioner-advocate, this lesson focuses on the workflow and decisions. As you might guess, the details of PAF can be quite complex. Some important notions to consider that may not be obvious from lesson, are these:

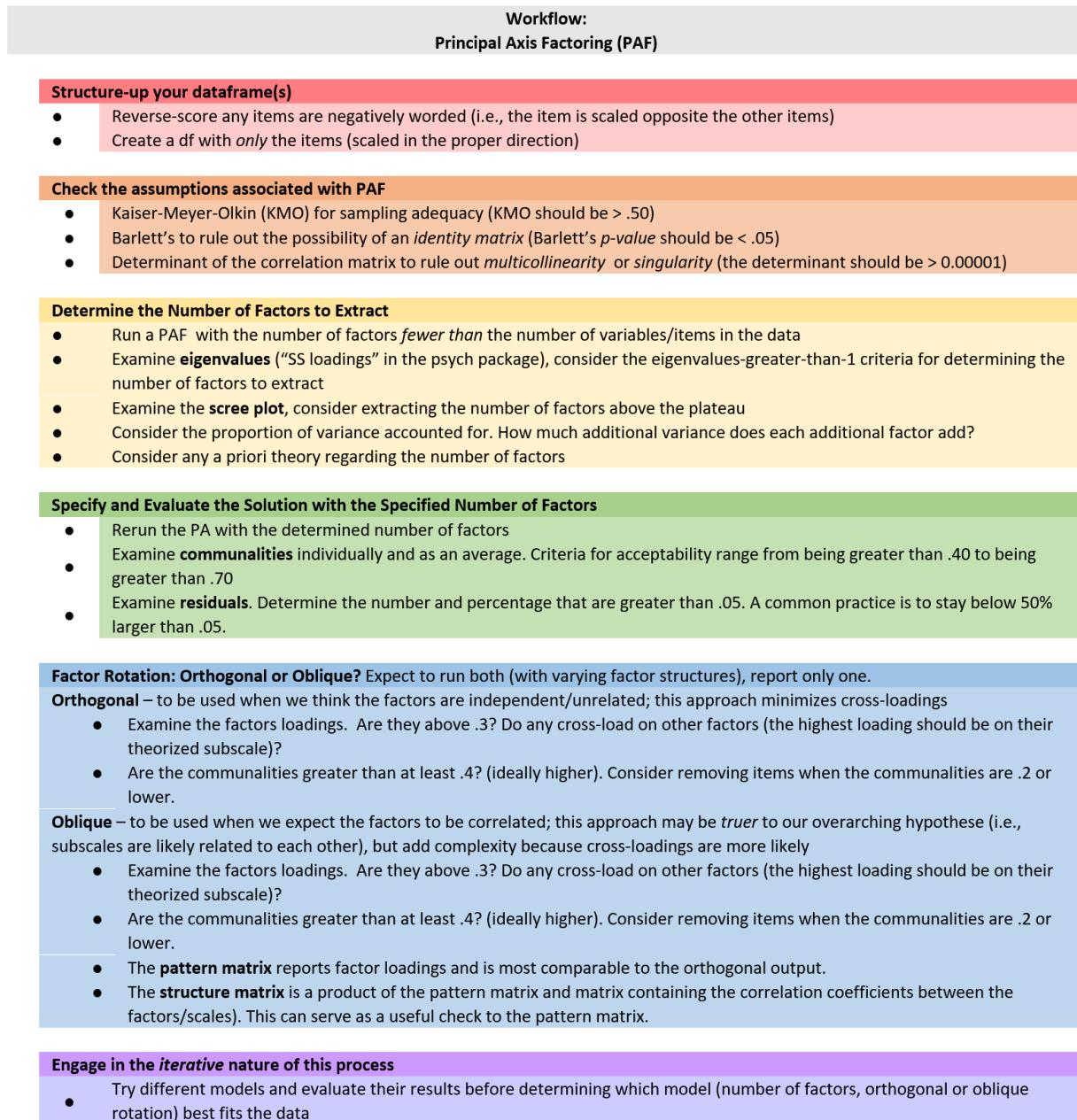


Figure 9.2: Image of the workflow for PAF

- The values of factor loadings are directly related to the correlation matrix.
  - Although I do not explain this in detail, nearly every analytic step attempts to convey this notion by presenting equivalent analytic options using the raw data and correlation matrix.
- PAF (like PCA and related EFA procedures) is about *dimension reduction* – our goal is fewer factors (think subscales) than there are items.
  - In this lesson's vignette there are 25 items on the scale and we will have 4 subscales.
- As a latent variable procedure, PAF is both *exploratory* and *factor analysis*. This is in contrast to our prior **PCA** lesson. Recall that PCA is a regression-based model and therefore not “factor analysis.”
- Matrix algebra (e.g., using the transpose of a matrix, multiplying matrices together) plays a critical role in the analytic solution.

## 9.4 Research Vignette

This lesson's research vignette emerges from Lewis and Neville's Gendered Racial Microaggressions Scale for Black Women [2015]. The article reports on two separate studies that comprised the development, refinement, and psychometric evaluation of two parallel versions (stress appraisal, frequency) of the scale. Below, I simulate data from the final construction of the stress appraisal version as the basis of the lecture. Items were on a 6-point Likert scale ranging from 0 (*not at all stressful*) to 5 (*extremely stressful*).

Lewis and Neville [2015] reported support for a total scale score (25 items) and four subscales. Below, I list the four subscales, their number of items, and a single example item. At the outset, let me provide a content advisory. For those who hold this particular identity (or related identities) the content in the items may be upsetting. In other lessons, I often provide a variable name that gives an indication of the primary content of the item. In the case of the GRMS, I will simply provide an abbreviation of the subscale name and its respective item number. This will allow us to easily inspect the alignment of the item with its intended factor, and hopefully minimize discomfort. If you are not a member of this particular identity, I encourage you to learn about these microaggressions by reading the article in its entirety. Please do not ask members of this group to explain why these microaggressions are harmful or ask if they have encountered them. The four factors, number of items, and sample item are as follows:

- Assumptions of Beauty and Sexual Objectification
  - 10 items
  - “Objectified me based on physical features.”
  - Abbreviated in the simulated data as “Obj#”
- Silenced and Marginalized
  - 7 items
  - “Someone has tried to ‘put me in my place.’”
  - Abbreviated in the simulated data as “Marg#”
- Strong Black Woman Stereotype

- 5 items
- “I have been told that I am too assertive.”
- Abbreviated in the simulated data as “Str#”
- Angry Black Woman Stereotype
  - 3 items
  - “Someone accused me of being angry when speaking calm.”
  - Abbreviated in the simulated data as “Ang#”

Three additional scales were reported in the Lewis and Neville article [2015]. Because (a) the focus of this lesson is on exploratory factor analytic approaches and, therefore, only requires item-level data for the scale, and (b) the article does not include correlations between the subscales/scales of all involved measures, I only simulated item-level data for the GRMS items.

Below, I walk through the data simulation. This is not an essential portion of the lesson, but I will lecture it in case you are interested. None of the items are negatively worded (relative to the other items), so there is no need to reverse-score any items.

Simulating the data involved using factor loadings, means, standard deviations, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

```
# Entering the intercorrelations, means, and standard deviations from
# the journal article

LewisGRMS_generating_model <- "
  #measurement model
  Objectification =~ .69*Obj1 + .69*Obj2 + .60*Obj3 + .59*Obj4 + .55*Obj5 + .55*Obj6 + .
  Marginalized =~ .93*Marg1 + .81*Marg2 + .69*Marg3 + .67*Marg4 + .61*Marg5 + .58*Marg6 + .
  Strong =~ .59*Str1 + .55*Str2 + .54*Str3 + .54*Str4 + .51*Str5
  Angry =~ .70*Ang1 + .69*Ang2 + .68*Ang3

  #Means
  Objectification ~ 1.85*1
  Marginalized ~ 2.67*1
  Strong ~ 1.61*1
  Angry ~ 2.29*1

  #Correlations
  Objectification ~~ .63*Marginalized
  Objectification ~~ .66*Strong
  Objectification ~~ .51*Angry

  Marginalized ~~ .59*Strong
  Marginalized ~~ .62*Angry

  Strong ~~ .61*Angry
```

```

"
set.seed(240311)
items <- lavaan::simulateData(model = LewisGRMS_generating_model, model.type = "sem",
  meanstructure = T, sample.nobs = 259, standardized = FALSE)

# used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(items))

# The code below loops through each column of the dataframe and
# assigns the scaling accordingly Rows 1 thru 26 are the GRMS items

for (i in 1:ncol(items)) {
  if (i >= 1 & i <= 26) {
    items[, i] <- scales::rescale(items[, i], c(1, 5))
  }
}

# rounding to integers so that the data resembles that which was
# collected
library(tidyverse)
items <- items %>%
  round(0)

# quick check of my work psych::describe(items)

```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do).

```

#write the simulated data as a .csv
#write.table(items, file="items.csv", sep=",", col.names=TRUE, row.names=FALSE)
#bring back the simulated dat from a .csv file
#items <- read.csv ("items.csv", header = TRUE)

#to save the df as an .rds (think "R object") file on your computer; it should save in the same
#saveRDS(items, "items.rds")
#bring back the simulated dat from an .rds file
#items <- readRDS("items.rds")

```

## 9.5 Working the Vignette

It may be useful to recall how we might understand factors in the psychometric sense:

- clusters of correlated items in an  $R$ -matrix
- statistical entities that can be plotted as classification axes where coordinates of variables along each axis represent the strength of the relationship between that variable to each factor.

- mathematical equations, resembling regression equations, where each variable is represented according to its relative weight

### 9.5.1 Data Prep

Since the first step is data preparation, let's start by:

- reverse coding any items that are phrased in the opposite direction
- creating a *df* (as an object) that only contains the items in their properly scored direction (i.e., you might need to replace the original item with the reverse-coded item); there should be no other variables (e.g., ID, demographic variables, other scales) in this df
  - because the GRMS has no items like this we can skip these two steps

Our example today requires no reverse coding and the dataset I simulated only has item-level data (with no ID and no other variables). This means we are ready to start the PAF process.

Let's take a look at (and make an object of) the correlation matrix.

```
GRMSr <- cor(items) #correlation matrix (with the negatively scored item already reversed) cr
round(GRMSr, 2)
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Marg1	Marg2	Marg3
Obj1	1.00	0.30	0.24	0.20	0.27	0.18	0.25	0.32	0.12	0.26	0.17	0.21	0.19
Obj2	0.30	1.00	0.32	0.24	0.27	0.21	0.24	0.29	0.26	0.19	0.08	0.19	0.14
Obj3	0.24	0.32	1.00	0.21	0.22	0.19	0.25	0.21	0.17	0.23	0.25	0.19	0.15
Obj4	0.20	0.24	0.21	1.00	0.36	0.19	0.27	0.27	0.23	0.26	0.16	0.13	0.17
Obj5	0.27	0.27	0.22	0.36	1.00	0.16	0.16	0.25	0.14	0.19	0.26	0.23	0.22
Obj6	0.18	0.21	0.19	0.19	0.16	1.00	0.16	0.19	0.14	0.10	0.16	0.06	0.05
Obj7	0.25	0.24	0.25	0.27	0.16	0.16	1.00	0.33	0.21	0.25	0.31	0.18	0.20
Obj8	0.32	0.29	0.21	0.27	0.25	0.19	0.33	1.00	0.16	0.26	0.12	0.10	0.12
Obj9	0.12	0.26	0.17	0.23	0.14	0.14	0.21	0.16	1.00	0.14	0.03	0.08	0.18
Obj10	0.26	0.19	0.23	0.26	0.19	0.10	0.25	0.26	0.14	1.00	0.10	0.10	0.20
Marg1	0.17	0.08	0.25	0.16	0.26	0.16	0.31	0.12	0.03	0.10	1.00	0.33	0.36
Marg2	0.21	0.19	0.19	0.13	0.23	0.06	0.18	0.10	0.08	0.10	0.33	1.00	0.35
Marg3	0.19	0.14	0.15	0.17	0.22	0.05	0.20	0.12	0.18	0.20	0.36	0.35	1.00
Marg4	0.21	0.15	0.20	0.24	0.21	0.13	0.21	0.17	0.07	0.17	0.41	0.20	0.37
Marg5	0.09	0.17	0.13	0.20	0.25	0.12	0.18	0.18	0.20	0.06	0.35	0.31	0.24
Marg6	0.22	0.21	0.11	0.22	0.24	0.22	0.31	0.20	0.12	0.14	0.34	0.28	0.31
Marg7	0.08	0.18	0.11	0.19	0.18	0.12	0.13	0.13	0.09	0.07	0.28	0.29	0.23
Str1	0.19	0.19	0.19	0.13	0.23	0.06	0.26	0.14	0.13	0.21	0.17	0.18	0.15
Str2	0.23	0.15	0.18	0.14	0.11	0.14	0.18	0.10	0.07	0.16	0.11	0.15	0.21
Str3	0.18	0.06	0.15	0.10	0.13	0.06	0.15	0.05	0.05	0.17	0.14	0.18	0.15
Str4	0.03	0.14	0.17	0.13	0.07	0.08	0.12	0.03	0.00	0.06	0.10	0.07	0.06
Str5	0.13	0.11	0.17	0.01	0.09	0.05	0.15	0.06	0.02	0.03	0.07	0.15	0.05
Ang1	0.06	0.01	0.15	0.14	0.11	0.04	0.25	0.08	0.12	0.06	0.21	0.19	0.13
Ang2	0.05	0.05	0.09	0.07	0.09	0.14	0.09	0.03	-0.01	0.13	0.13	0.21	0.14

Ang3	0.21	0.10	0.18	0.19	0.11	0.11	0.23	0.08	0.08	0.14	0.25	0.20	0.14
Marg4		Marg5	Marg6	Marg7	Str1	Str2	Str3	Str4	Str5	Ang1	Ang2	Ang3	
Obj1	0.21	0.09	0.22	0.08	0.19	0.23	0.18	0.03	0.13	0.06	0.05	0.21	
Obj2	0.15	0.17	0.21	0.18	0.19	0.15	0.06	0.14	0.11	0.01	0.05	0.10	
Obj3	0.20	0.13	0.11	0.11	0.19	0.18	0.15	0.17	0.17	0.15	0.09	0.18	
Obj4	0.24	0.20	0.22	0.19	0.13	0.14	0.10	0.13	0.01	0.14	0.07	0.19	
Obj5	0.21	0.25	0.24	0.18	0.23	0.11	0.13	0.07	0.09	0.11	0.09	0.11	
Obj6	0.13	0.12	0.22	0.12	0.06	0.14	0.06	0.08	0.05	0.04	0.14	0.11	
Obj7	0.21	0.18	0.31	0.13	0.26	0.18	0.15	0.12	0.15	0.25	0.09	0.23	
Obj8	0.17	0.18	0.20	0.13	0.14	0.10	0.05	0.03	0.06	0.08	0.03	0.08	
Obj9	0.07	0.20	0.12	0.09	0.13	0.07	0.05	0.00	0.02	0.12	-0.01	0.08	
Obj10	0.17	0.06	0.14	0.07	0.21	0.16	0.17	0.06	0.03	0.06	0.13	0.14	
Marg1	0.41	0.35	0.34	0.28	0.17	0.11	0.14	0.10	0.07	0.21	0.13	0.25	
Marg2	0.20	0.31	0.28	0.29	0.18	0.15	0.18	0.07	0.15	0.19	0.21	0.20	
Marg3	0.37	0.24	0.31	0.23	0.15	0.21	0.15	0.06	0.05	0.13	0.14	0.14	
Marg4	1.00	0.27	0.28	0.24	0.13	0.17	0.13	0.16	-0.01	0.11	0.17	0.20	
Marg5	0.27	1.00	0.27	0.23	0.13	0.06	0.20	0.11	0.08	0.04	0.10	0.22	
Marg6	0.28	0.27	1.00	0.26	0.12	0.28	0.17	0.14	0.09	0.13	0.21	0.16	
Marg7	0.24	0.23	0.26	1.00	0.12	-0.01	0.05	0.05	0.03	0.18	0.12	0.08	
Str1	0.13	0.13	0.12	0.12	1.00	0.16	0.22	0.14	0.18	0.18	0.05	0.06	
Str2	0.17	0.06	0.28	-0.01	0.16	1.00	0.19	0.17	0.18	0.11	0.16	0.12	
Str3	0.13	0.20	0.17	0.05	0.22	0.19	1.00	0.27	0.19	0.27	0.13	0.22	
Str4	0.16	0.11	0.14	0.05	0.14	0.17	0.27	1.00	0.11	0.12	0.04	0.04	
Str5	-0.01	0.08	0.09	0.03	0.18	0.18	0.19	0.11	1.00	0.15	0.11	0.12	
Ang1	0.11	0.04	0.13	0.18	0.18	0.11	0.27	0.12	0.15	1.00	0.23	0.26	
Ang2	0.17	0.10	0.21	0.12	0.05	0.16	0.13	0.04	0.11	0.23	1.00	0.27	
Ang3	0.20	0.22	0.16	0.08	0.06	0.12	0.22	0.04	0.12	0.26	0.27	1.00	

In case you want to examine it in sections (easier to view):

```
# round(GRMSr[, 1:8], 2) round(GRMSr[, 9:16], 2) round(GRMSr[, 17:25],
# 2)
```

As with PCA, we can analyze the data with either raw data or correlation matrix. I will do both to demonstrate (a) that it's possible and to (b) continue emphasizing that this is a *structural* analysis. That is, we are trying to see if our more parsimonious extraction *reproduces* this original correlation matrix.

### 9.5.1.1 Three Diagnostic Tests to Evaluate the Appropriateness of the Data for Component (or Factor)Analysis

#### 9.5.1.2 Is my sample adequate for PAF?

We return to the **KMO** (Kaiser-Meyer-Olkin), an index of *sampling adequacy* that can be used with the actual sample to let us know if the sample size is sufficient (or if we should collect more data).

Kaiser's 1974 recommendations were:

- bare minimum of .5
- values between .5 and .7 are mediocre
- values between .7 and .8 are good
- values above .9 are superb

We use the *KMO()* function from the *psych* package with either raw or matrix dat.

```
psych::KMO(items)
```

Kaiser-Meyer-Olkin factor adequacy

Call: psych::KMO(r = items)

Overall MSA = 0.84

MSA for each item =

Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Marg1	Marg2	Marg3
0.85	0.85	0.88	0.86	0.87	0.86	0.87	0.85	0.76	0.85	0.83	0.87	0.87
Marg4	Marg5	Marg6	Marg7	Str1	Str2	Str3	Str4	Str5	Ang1	Ang2	Ang3	
0.87	0.82	0.88	0.84	0.87	0.84	0.79	0.74	0.81	0.74	0.75	0.82	

```
# psych::KMO(GRMSr) #for the KMO function, do not specify sample size
# if using the matrix form of the data
```

We examine the KMO values for both the overall matrix and the individual items.

At the matrix level, our *KMO* = .84, which falls in between Kaiser's definitions of *good* and *superb*.

At the item level, the KMO should be > .50. Variables with values below .5 should be evaluated for exclusion from the analysis (or run the analysis with and without the variable and compare the difference). Because removing/adding variables impacts the KMO, be sure to re-evaluate.

At the item level, our KMO values range between .71 (Ang1, Ang2) and .88 (Obj3, Marg6).

Considering both item- and matrix- levels, we conclude that the sample size and the data are adequate for component (or factor) analysis.

### 9.5.1.3 Are there correlations among the variables that are big enough to be analyzed?

**Bartlett's** lets us know if a matrix is an *identity matrix*. In an identity matrix all correlation coefficients (everything on the off-diagonal) would be 0.0 (and everything on the diagonal would be 1.0).

A significant Barlett's (i.e.,  $p < .05$ ) tells that the *R*-matrix is not an identity matrix. That is, there are some relationships between variables that can be analyzed.

The *cortest.bartlett()* function in the *psych* package and can be run either from the raw data or R matrix formats.

```
psych::cortest.bartlett(items) #from the raw data
```

R was not square, finding R from data

Our Bartlett's test is significant:  $\chi^2(300) = 1113.30, p < .001$ . This supports a component (or factor) analytic approach for investigating the data.

#### 9.5.1.4 Is there multicollinearity or singularity in my data?

The **determinant of the correlation matrix** should be greater than 0.00001 (that would be 4 zeros before the 1). If it is smaller than 0.00001 then we may have an issue with *multicollinearity* (i.e., variables that are too highly correlated) or *singularity* (variables that are perfectly correlated).

The determinant function comes from base R. It is easiest to compute when the correlation matrix is the object. However, it is also possible to specify the command to work with the raw data.

```
# det(GRMSr)
det(cor(items)) #if using the raw data
```

[1] 0.01140074

With a value of 0.000115, our determinant is greater than the 0.00001 requirement. If it were not, then we could identify problematic variables (i.e., those correlating too highly with others and those not correlating sufficiently with others) and re-run the diagnostic statistics.

### 9.5.1.5 APA Style Summary So Far

Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00 – values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was .84, indicating acceptable sampling adequacy. The Barlett's Test of Sphericity examines

whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(300) = 1113.30, p < .001$ , indicating the correlations between items are sufficiently large enough for principal axis factoring. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.01140 and, again, indicated that our data was suitable for the analysis.

*Note:* If this looks familiar, it is! The same diagnostics are used in PAF and **PCA**.

### 9.5.2 Principal Axis Factoring (PAF)

We can use the `fa()` function, specifying `fm = "pa"` from the `psych` package with raw or matrix data.

One difference from PCA is that factor analysis will not (cannot) calculate as many factors as there are items. This means that we should select a reasonable number, like 20 (since there are 25 items). However, I received a number of errors/warnings and 13 is the first number that would run. I also received the warning, "maximum iteration exceeded." Therefore I increased "max.iter" to 100.

Our goal is to begin to get an idea of the cumulative variance explained and number of factors to extract. If we think there are four factors, we simply need to specify more than four factors on the `nfactors = ##` command. As long as that number is less than the total number of items, it does not matter what that number is.

```
# grmsPAF1 <- psych::fa(GRMSr, nfactors=10, fm = 'pa', max.iter =
# 100, rotate='none')# using the matrix data and specifying the # of
# factors.

grmsPAF1 <- psych::fa(items, nfactors = 13, fm = "pa", max.iter = 100,
rotate = "none") # using raw data and specifying the max number of factors
```

maximum iteration exceeded

```
# I received the warning 'maximum iteration exceeded'. It gave
# output, but it's best if we don't get that warning, so I increased
# it to 100.
```

grmsPAF1 #this object holds a great deal of information

```
Factor Analysis using method = pa
Call: psych::fa(r = items, nfactors = 13, rotate = "none", max.iter = 100,
fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
    PA1    PA2    PA3    PA4    PA5    PA6    PA7    PA8    PA9    PA10   PA11   PA12
Obj1  0.50  0.37  0.08 -0.33 -0.27 -0.03 -0.29 -0.10  0.01  0.06 -0.02  0.18
Obj2  0.45  0.34 -0.08  0.04  0.10  0.12 -0.06  0.11  0.13  0.14  0.14 -0.05
```



Ang3 0.07 0.37 0.63 4.5

	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9	PA10	PA11
SS loadings	4.54	1.17	0.97	0.64	0.60	0.56	0.47	0.47	0.44	0.42	0.35
Proportion Var	0.18	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Cumulative Var	0.18	0.23	0.27	0.29	0.32	0.34	0.36	0.38	0.39	0.41	0.43
Proportion Explained	0.40	0.10	0.09	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.03
Cumulative Proportion	0.40	0.51	0.60	0.65	0.71	0.76	0.80	0.84	0.88	0.92	0.95
	PA12	PA13									
SS loadings	0.31	0.27									
Proportion Var	0.01	0.01									
Cumulative Var	0.44	0.45									
Proportion Explained	0.03	0.02									
Cumulative Proportion	0.98	1.00									

Mean item complexity = 4.1

Test of the hypothesis that 13 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
df of the model are 53 and the objective function was 0.09

The root mean square of the residuals (RMSR) is 0.01

The df corrected root mean square of the residuals is 0.02

The harmonic n.obs is 259 with the empirical chi square 16.54 with prob < 1  
The total n.obs was 259 with Likelihood Chi Square = 22.09 with prob < 1

Tucker Lewis Index of factoring reliability = 1.226

RMSEA index = 0 and the 90 % confidence intervals are 0 0

BIC = -272.42

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	PA1	PA2	PA3	PA4	PA5	PA6
Correlation of (regression) scores with factors	0.95	0.85	0.82	0.81	0.79	0.73
Multiple R square of scores with factors	0.91	0.72	0.67	0.66	0.63	0.54
Minimum correlation of possible factor scores	0.81	0.44	0.34	0.32	0.26	0.07
	PA7	PA8	PA9	PA10	PA11	
Correlation of (regression) scores with factors	0.72	0.68	0.68	0.67	0.63	
Multiple R square of scores with factors	0.51	0.47	0.46	0.45	0.39	
Minimum correlation of possible factor scores	0.03	-0.06	-0.09	-0.10	-0.21	
	PA12	PA13				
Correlation of (regression) scores with factors	0.62	0.58				
Multiple R square of scores with factors	0.38	0.34				
Minimum correlation of possible factor scores	-0.24	-0.33				

*The total variance for a particular variable will have two factors: some variance will be shared with other variables (common variance) and some variance will be specific to that measure (unique variance). Random variance is also specific to one item, but not reliably so.*

We can examine this most easily by examining the matrix (second screen).

The columns PA1 thru PA10 are the (uninteresting at this point) unrotated loadings. These are the loading from each factor to each variable. PA stands for “principal axis.”

Scrolling to the far right we are interested in:

**Communalities** are represented as  $h^2$ . These are the proportions of common variance present in the variables. A variable that has no specific (or random) variance would have a communality of 1.0. If a variable shares none of its variance with any other variable, its communality would be 0.0. As a point of comparison, in PCA these started as 1.0 because we extracted the same number of components as items. In PAF, because we must extract fewer factors than items, these will have unique values.

\*\*Uniquenesses\* are represented as  $u^2$ . These are the amount of unique variance for each variable. They are calculated as  $1 - h^2$  (or 1 minus the communality).

The final column, *com* represents *item complexity*. This is an indication of how well an item reflects a single construct. If it is 1.0 then the item loads only on one component, if it is 2.0, it loads evenly on two components, and so forth. For now, we can ignore this. *I mostly wanted to reassure you that “com” is not “communality” –  $h^2$  is communality.*

Let’s switch to the first screen of output.

**Eigenvalues** are displayed in the row called, *SS loadings* (i.e., the sum of squared loadings). They represent the variance explained by the particular linear component. PA1 explains 4.54 units of variance (out of a possible 25; the # of potential factors). As a proportion, this is  $4.54/25 = 0.1816$  (reported in the *Proportion Var* row).

```
4.54/25
```

```
[1] 0.1816
```

We inspect the eigenvalues to see how many are  $> 1.0$  (Kaiser’s eigenvalue  $> 1$  criteria criteria). We see there are two that meet Kaiser’s criteria and three that meet Joliffe’s criteria (eigenvalues  $> .77$ ).

**Cumulative Var** is helpful to determine how many factors we’d like to retain to balance parsimony (few as possible) with the amount of variance we want to explain. The eigenvalues are in descending order. Using both Kaiser’s criteria (eigenvalue  $> 1.0$ ), Joliffe’s criteria (eigenvalue  $> 0.7$ ) criteria, and the a priori theory related to the Lewis and Neville [2015] article, we landed on a four-factor solution. Extracting four factors (like we did with PCA will) will explain 29% of the variance. Eigenvalues are only one criteria, let’s look at the scree plot.

**Scree plot:**

Eigenvalues are stored in the *grmsPAF1* object’s variable, “values”. We can see all the values captured by this object with the *names()* function:

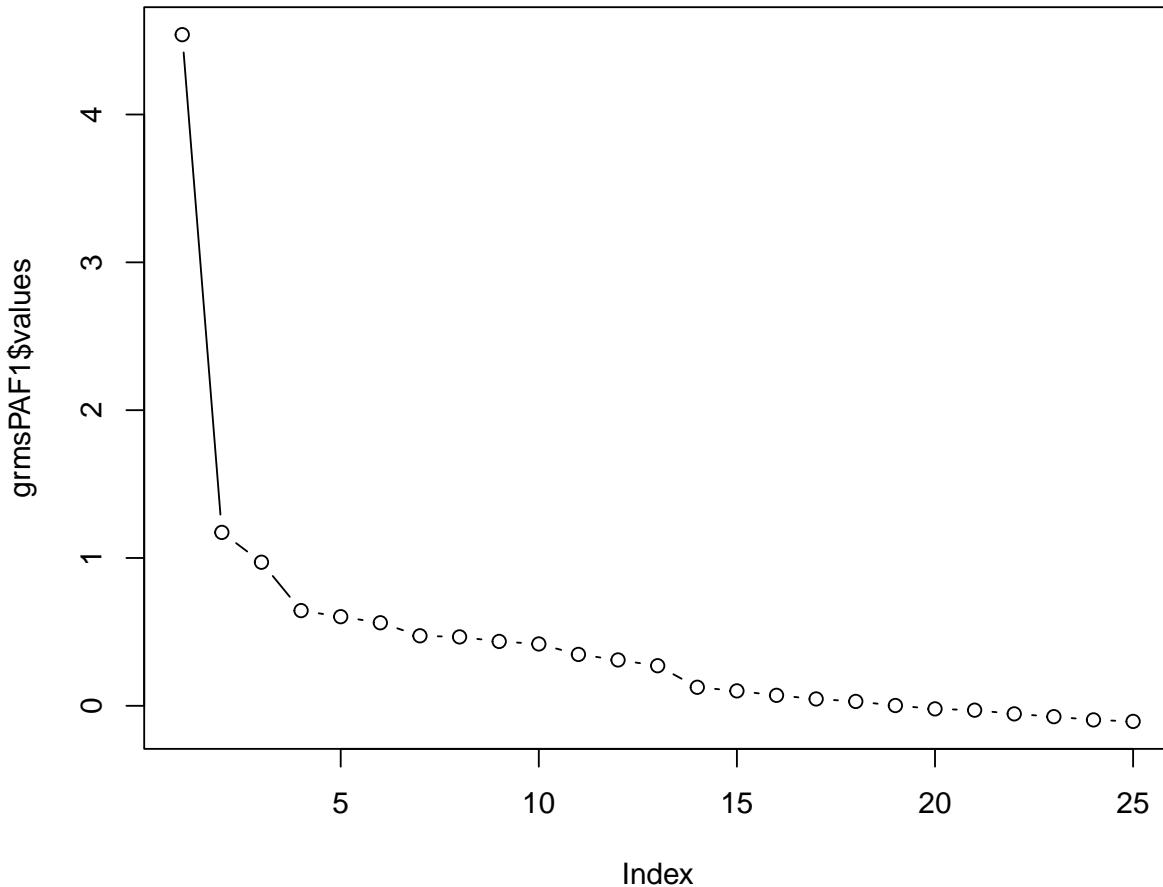
```
names(grmsPAF1)
```

```
[1] "residual"           "dof"                  "chi"
```

```
[4] "nh"                      "rms"                      "EPVAL"
[7] "crms"                    "EBIC"                     "ESABIC"
[10] "fit"                     "fit.off"                   "sd"
[13] "factors"                 "complexity"                "n.obs"
[16] "objective"               "criteria"                  "STATISTIC"
[19] "PVAL"                    "Call"                     "null.model"
[22] "null.dof"                "null.chisq"                "TLI"
[25] "F0"                      "RMSEA"                    "BIC"
[28] "SABIC"                   "r.scores"                  "R2"
[31] "valid"                   "score.cor"                 "weights"
[34] "rotation"                 "hyperplane"                "communality"
[37] "communalities"            "uniquenesses"              "values"
[40] "e.values"                 "loadings"                  "model"
[43] "fm"                      "Structure"                 "communality.iterations"
[46] "method"                   "scores"                    "R2.scores"
[49] "r"                        "np.obs"                    "fn"
[52] "Vaccounted"
```

Plotting the *eigenvalues* produces a scree plot. We can use this to further guage the number of factors we should extract.

```
plot(grmsPAF1$values, type = "b") #type = "b" gives us "both" lines and points; type = "l" gi
```



We look for the point of *inflection*. That is, where the baseline levels out into a plateau. I can see inflections after 1, 2, 3, and 4.

#### 9.5.2.1 Specifying the number of factors

Having determined the number of components, we must rerun the analysis with this specification. Especially when researchers may not have a clear theoretical structure that guides the process, researchers may do this iteratively with varying numbers of factors. Lewis and Neville [Lewis and Neville, 2015] examined solutions with 2, 3, 4, and 5 factors (they conducted a parallel *factor* analysis; in contrast this lesson demonstrates principal axis factoring).

```
# grmsPAF2 <- psych::fa(GRMSr, nfactors=4, fm = 'pa', rotate='none')
grmsPAF2 <- psych::fa(items, nfactors = 4, fm = "pa", rotate = "none") #can copy prior script
grmsPAF2
```

Factor Analysis using method = pa

Call: psych::fa(r = items, nfactors = 4, rotate = "none", fm = "pa")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	PA1	PA2	PA3	PA4	h2	u2	com
Obj1	0.46	0.24	0.05	-0.06	0.28	0.72	1.6
Obj2	0.45	0.36	-0.06	0.12	0.35	0.65	2.1
Obj3	0.45	0.16	0.11	0.04	0.24	0.76	1.4
Obj4	0.47	0.21	-0.11	-0.09	0.28	0.72	1.6
Obj5	0.48	0.12	-0.13	0.10	0.27	0.73	1.4
Obj6	0.31	0.14	-0.05	-0.13	0.13	0.87	1.8
Obj7	0.53	0.12	0.07	-0.08	0.30	0.70	1.2
Obj8	0.42	0.35	-0.11	-0.09	0.32	0.68	2.2
Obj9	0.29	0.21	-0.07	0.06	0.14	0.86	2.1
Obj10	0.38	0.23	0.07	-0.13	0.22	0.78	2.0
Marg1	0.54	-0.34	-0.19	0.01	0.44	0.56	2.0
Marg2	0.48	-0.25	-0.05	0.14	0.32	0.68	1.7
Marg3	0.49	-0.20	-0.16	0.04	0.30	0.70	1.6
Marg4	0.50	-0.17	-0.19	-0.09	0.32	0.68	1.6
Marg5	0.45	-0.16	-0.19	0.15	0.29	0.71	1.9
Marg6	0.53	-0.12	-0.10	-0.07	0.31	0.69	1.2
Marg7	0.38	-0.17	-0.24	0.08	0.23	0.77	2.3
Str1	0.38	0.08	0.19	0.23	0.24	0.76	2.3
Str2	0.35	0.03	0.23	-0.06	0.18	0.82	1.8
Str3	0.36	-0.14	0.39	0.12	0.32	0.68	2.5
Str4	0.24	-0.03	0.21	0.15	0.13	0.87	2.7
Str5	0.23	-0.02	0.31	0.14	0.17	0.83	2.3
Ang1	0.33	-0.19	0.27	-0.07	0.23	0.77	2.7
Ang2	0.29	-0.23	0.15	-0.27	0.23	0.77	3.5
Ang3	0.39	-0.16	0.16	-0.25	0.26	0.74	2.5

	PA1	PA2	PA3	PA4
SS loadings	4.34	0.98	0.78	0.40
Proportion Var	0.17	0.04	0.03	0.02
Cumulative Var	0.17	0.21	0.24	0.26
Proportion Explained	0.67	0.15	0.12	0.06
Cumulative Proportion	0.67	0.82	0.94	1.00

Mean item complexity = 2

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

```
Tucker Lewis Index of factoring reliability = 1.008
RMSEA index = 0 and the 90 % confidence intervals are 0 0.025
BIC = -943.32
Fit based upon off diagonal values = 0.96
Measures of factor score adequacy
```

	PA1	PA2	PA3	PA4
Correlation of (regression) scores with factors	0.93	0.77	0.72	0.59
Multiple R square of scores with factors	0.86	0.59	0.51	0.35
Minimum correlation of possible factor scores	0.72	0.17	0.03	-0.30

Our eigenvalues/SS loadings wiggle around a bit from the initial run. With four factors, we now, cumulatively, explain 26% of the variance.

*Communality* is the proportion of common variance within a variable. Changing from 13 to 4 factors changed these values ( $h^2$ ) as well as their associated *uniquenesses* ( $u^2$ ), which are calculated as “1.0 minus the communality.”

Now we see that 28% of the variance associated with Obj1 is common/shared (the  $h^2$  value).

As a reminder of what we are doing, recall that we are looking for a more *parsimonious* explanation than 25 items on the GRMS. By respecifying a smaller number of factors, we lose some information. That is, the retained factors (now 4) cannot explain all of the variance present in the data (as we saw, it explains about 28%, cumulatively). The amount of variance explained in each variable is represented by the communalities after extraction.

We can also inspect the communalities through the lens of Kaiser's criterion (the eigenvalue  $> 1$  criteria) to see if we think that four was a good number of factors to extract.

Kaiser's criterion is believed to be accurate if:

- when there are fewer than 30 variables (we had 25) and, after extraction, the communalities are greater than .70
  - looking at our data, none of the communalities is  $> .70$ , so, this does not support extracting four components
- when the sample size is greater than 250 (ours was 259) and the average communality is  $> .60$ 
  - again, our communalities were lower than this

Using the *names()* function again, we see that “communality” is available for manipulation.

```
names(grmsPAF2)
```

[1] "residual"	"dof"	"chi"
[4] "nh"	"rms"	"EPVAL"
[7] "crms"	"EBIC"	"ESABIC"
[10] "fit"	"fit.off"	"sd"
[13] "factors"	"complexity"	"n.obs"
[16] "objective"	"criteria"	"STATISTIC"

```
[19] "PVAL"           "Call"           "null.model"
[22] "null.dof"       "null.chisq"      "TLI"
[25] "FO"              "RMSEA"          "BIC"
[28] "SABIC"          "r.scores"        "R2"
[31] "valid"          "score.cor"       "weights"
[34] "rotation"        "hyperplane"      "communality"
[37] "communalities"   "uniquenesses"    "values"
[40] "e.values"        "loadings"         "model"
[43] "fm"              "Structure"       "communality.iterations"
[46] "method"          "scores"          "R2.scores"
[49] "r"               "np.obs"          "fn"
[52] "Vaccounted"
```

We can use this value to calculate their mean.

```
mean(grmsPAF2$communality)
```

```
[1] 0.2599292
```

```
# sum(grmsPAF2$communality) #
```

We see that our average communality is 0.26. These two criteria suggest that we may not have the best solution. That said (in our defense):

- We used the scree plot as a guide and it was very clear.
- We have an adequate sample size and that was supported with the KMO.
- Are the number of factors consistent with theory? We have not yet inspected the factor loadings. This will provide us with more information.

We could do several things:

- rerun with a different number of factors (recall Lewis and Neville [2015] ran models with 2, 3, 4, and 5 factors)
- conduct more diagnostics tests
  - reproduced correlation matrix
  - the difference between the reproduced correlation matrix and the correlation matrix in the data

The *factor.model()* function in *psych* produces the *reproduced correlation matrix* by using the *loadings* in our extracted object. Conceptually, this matrix is the correlations that should be produced if we did not have the raw data but we only had the factor loadings. We could do fancy matrix algebra and produce these.

The questions, though, is: How close did we get? How different is the *reproduced correlation matrix* from *GRMSmatrix* – the *R*-matrix produced from our raw data.

```
round(psych::factor.model(grmsPAF2$loadings), 3) #produces the reproduced correlation matrix
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Marg1	Marg2
Obj1	0.275	0.282	0.250	0.263	0.238	0.181	0.279	0.275	0.177	0.241	0.157	0.149
Obj2	0.282	0.350	0.260	0.278	0.282	0.176	0.265	0.308	0.217	0.232	0.135	0.146
Obj3	0.250	0.260	0.245	0.229	0.228	0.153	0.262	0.230	0.160	0.212	0.170	0.177
Obj4	0.263	0.278	0.229	0.278	0.255	0.189	0.268	0.286	0.181	0.227	0.202	0.165
Obj5	0.238	0.282	0.228	0.255	0.275	0.160	0.250	0.250	0.181	0.189	0.244	0.221
Obj6	0.181	0.176	0.153	0.189	0.160	0.133	0.186	0.194	0.115	0.162	0.130	0.100
Obj7	0.279	0.265	0.262	0.268	0.250	0.186	0.302	0.261	0.168	0.243	0.230	0.208
Obj8	0.275	0.308	0.230	0.286	0.250	0.194	0.261	0.316	0.197	0.242	0.129	0.107
Obj9	0.177	0.217	0.160	0.181	0.181	0.115	0.168	0.197	0.136	0.146	0.100	0.098
Obj10	0.241	0.232	0.212	0.227	0.189	0.162	0.243	0.242	0.146	0.219	0.115	0.105
Marg1	0.157	0.135	0.170	0.202	0.244	0.130	0.230	0.129	0.100	0.115	0.444	0.357
Marg2	0.149	0.146	0.177	0.165	0.221	0.100	0.208	0.107	0.098	0.105	0.357	0.318
Marg3	0.166	0.163	0.173	0.200	0.235	0.127	0.218	0.149	0.113	0.125	0.362	0.298
Marg4	0.183	0.163	0.173	0.224	0.234	0.151	0.234	0.177	0.116	0.148	0.361	0.279
Marg5	0.149	0.175	0.163	0.184	0.238	0.108	0.192	0.140	0.119	0.102	0.338	0.290
Marg6	0.214	0.194	0.207	0.239	0.246	0.162	0.262	0.198	0.132	0.177	0.345	0.280
Marg7	0.114	0.134	0.120	0.159	0.200	0.095	0.153	0.119	0.095	0.077	0.306	0.246
Str1	0.190	0.217	0.215	0.154	0.192	0.092	0.205	0.146	0.127	0.147	0.146	0.187
Str2	0.185	0.148	0.187	0.151	0.138	0.110	0.210	0.138	0.090	0.165	0.139	0.144
Str3	0.146	0.104	0.188	0.087	0.118	0.060	0.192	0.049	0.055	0.118	0.172	0.209
Str4	0.107	0.106	0.134	0.073	0.102	0.044	0.128	0.056	0.059	0.082	0.104	0.135
Str5	0.108	0.092	0.139	0.056	0.080	0.036	0.130	0.041	0.047	0.087	0.074	0.120
Ang1	0.125	0.055	0.146	0.093	0.094	0.074	0.177	0.049	0.033	0.111	0.197	0.188
Ang2	0.100	0.002	0.097	0.093	0.061	0.085	0.154	0.047	0.008	0.102	0.202	0.150
Ang3	0.164	0.077	0.157	0.152	0.121	0.123	0.216	0.112	0.054	0.155	0.233	0.185
	Marg3	Marg4	Marg5	Marg6	Marg7	Str1	Str2	Str3	Str4	Str5	Ang1	Ang2
Obj1	0.166	0.183	0.149	0.214	0.114	0.190	0.185	0.146	0.107	0.108	0.125	0.100
Obj2	0.163	0.163	0.175	0.194	0.134	0.217	0.148	0.104	0.106	0.092	0.055	0.002
Obj3	0.173	0.173	0.163	0.207	0.120	0.215	0.187	0.188	0.134	0.139	0.146	0.097
Obj4	0.200	0.224	0.184	0.239	0.159	0.154	0.151	0.087	0.073	0.056	0.093	0.093
Obj5	0.235	0.234	0.238	0.246	0.200	0.192	0.138	0.118	0.102	0.080	0.094	0.061
Obj6	0.127	0.151	0.108	0.162	0.095	0.092	0.110	0.060	0.044	0.036	0.074	0.085
Obj7	0.218	0.234	0.192	0.262	0.153	0.205	0.210	0.192	0.128	0.130	0.177	0.154
Obj8	0.149	0.177	0.140	0.198	0.119	0.146	0.138	0.049	0.056	0.041	0.049	0.047
Obj9	0.113	0.116	0.119	0.132	0.095	0.127	0.090	0.055	0.059	0.047	0.033	0.008
Obj10	0.125	0.148	0.102	0.177	0.077	0.147	0.165	0.118	0.082	0.087	0.111	0.102
Marg1	0.362	0.361	0.338	0.345	0.306	0.146	0.139	0.172	0.104	0.074	0.197	0.202
Marg2	0.298	0.279	0.290	0.280	0.246	0.187	0.144	0.209	0.135	0.120	0.188	0.150
Marg3	0.304	0.303	0.290	0.295	0.258	0.149	0.129	0.148	0.097	0.072	0.157	0.151
Marg4	0.303	0.318	0.276	0.307	0.253	0.121	0.133	0.120	0.074	0.047	0.155	0.176
Marg5	0.290	0.276	0.291	0.268	0.256	0.158	0.103	0.131	0.098	0.068	0.122	0.097
Marg6	0.295	0.307	0.268	0.308	0.237	0.158	0.165	0.161	0.102	0.084	0.178	0.182
Marg7	0.258	0.253	0.256	0.237	0.234	0.103	0.068	0.075	0.058	0.025	0.088	0.087
Str1	0.149	0.121	0.158	0.158	0.103	0.239	0.166	0.228	0.164	0.175	0.146	0.055

```

Str2 0.129 0.133 0.103 0.165 0.068 0.166 0.180 0.205 0.123 0.143 0.177 0.143
Str3 0.148 0.120 0.131 0.161 0.075 0.228 0.205 0.318 0.191 0.225 0.244 0.160
Str4 0.097 0.074 0.098 0.102 0.058 0.164 0.123 0.191 0.125 0.141 0.131 0.065
Str5 0.072 0.047 0.068 0.084 0.025 0.175 0.143 0.225 0.141 0.169 0.155 0.080
Ang1 0.157 0.155 0.122 0.178 0.088 0.146 0.177 0.244 0.131 0.155 0.226 0.199
Ang2 0.151 0.176 0.097 0.182 0.087 0.055 0.143 0.160 0.065 0.080 0.199 0.230
Ang3 0.187 0.211 0.134 0.225 0.113 0.108 0.183 0.195 0.095 0.109 0.220 0.238
      Ang3
Obj1 0.164
Obj2 0.077
Obj3 0.157
Obj4 0.152
Obj5 0.121
Obj6 0.123
Obj7 0.216
Obj8 0.112
Obj9 0.054
Obj10 0.155
Marg1 0.233
Marg2 0.185
Marg3 0.187
Marg4 0.211
Marg5 0.134
Marg6 0.225
Marg7 0.113
Str1 0.108
Str2 0.183
Str3 0.195
Str4 0.095
Str5 0.109
Ang1 0.220
Ang2 0.238
Ang3 0.263

```

We're not really interested in this matrix. We just need it to compare it to the *GRMSmatrix* to produce the residuals. We do that next.

**Residuals** are the difference between the reproduced (i.e., those created from our factor loadings) and *R*-matrix produced by the raw data.

If we look at the  $r_{Obj1Obj2}$  in our original correlation matrix (theoretically from the raw data [although we simulated data]), the value is 0.30. The reproduced correlation for this pair is 0.282. The difference is 0.018. The residuals table below shows 0.020 (rounding error).

```
.30 - .282
```

```
[1] 0.018
```

By using the `factor.residuals()` function we can calculate the residuals. Here we will see this difference calculated for us, for all the elements in the matrix.

```
round(psych::factor.residuals(GRMSr, grmsPAF2$loadings), 3)
```

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10
Obj1	0.725	0.020	-0.009	-0.066	0.031	-0.005	-0.026	0.041	-0.061	0.017
Obj2	0.020	0.650	0.056	-0.034	-0.010	0.033	-0.023	-0.023	0.046	-0.041
Obj3	-0.009	0.056	0.755	-0.023	-0.008	0.036	-0.010	-0.019	0.010	0.018
Obj4	-0.066	-0.034	-0.023	0.722	0.103	-0.002	0.005	-0.012	0.045	0.032
Obj5	0.031	-0.010	-0.008	0.103	0.725	-0.005	-0.085	-0.001	-0.039	-0.003
Obj6	-0.005	0.033	0.036	-0.002	-0.005	0.867	-0.029	-0.004	0.024	-0.060
Obj7	-0.026	-0.023	-0.010	0.005	-0.085	-0.029	0.698	0.067	0.039	0.007
Obj8	0.041	-0.023	-0.019	-0.012	-0.001	-0.004	0.067	0.684	-0.039	0.014
Obj9	-0.061	0.046	0.010	0.045	-0.039	0.024	0.039	-0.039	0.864	-0.010
Obj10	0.017	-0.041	0.018	0.032	-0.003	-0.060	0.007	0.014	-0.010	0.781
Marg1	0.011	-0.056	0.081	-0.046	0.012	0.029	0.077	-0.005	-0.070	-0.019
Marg2	0.057	0.046	0.011	-0.037	0.010	-0.044	-0.029	-0.007	-0.021	-0.003
Marg3	0.027	-0.023	-0.020	-0.030	-0.012	-0.073	-0.016	-0.031	0.063	0.077
Marg4	0.023	-0.011	0.030	0.015	-0.020	-0.019	-0.023	-0.006	-0.046	0.021
Marg5	-0.057	-0.001	-0.031	0.019	0.012	0.017	-0.010	0.044	0.077	-0.045
Marg6	0.009	0.016	-0.093	-0.016	-0.002	0.057	0.052	-0.001	-0.013	-0.036
Marg7	-0.034	0.047	-0.011	0.029	-0.019	0.028	-0.021	0.014	-0.003	-0.004
Str1	0.000	-0.027	-0.028	-0.020	0.042	-0.030	0.057	-0.010	0.005	0.060
Str2	0.043	0.002	-0.004	-0.012	-0.026	0.029	-0.028	-0.034	-0.022	-0.001
Str3	0.031	-0.042	-0.033	0.015	0.008	-0.004	-0.042	-0.001	-0.004	0.050
Str4	-0.072	0.037	0.039	0.062	-0.036	0.038	-0.006	-0.024	-0.057	-0.026
Str5	0.019	0.021	0.027	-0.048	0.014	0.015	0.018	0.019	-0.025	-0.061
Ang1	-0.065	-0.040	0.002	0.050	0.013	-0.034	0.068	0.034	0.091	-0.051
Ang2	-0.054	0.045	-0.007	-0.018	0.025	0.060	-0.065	-0.014	-0.022	0.031
Ang3	0.047	0.027	0.023	0.037	-0.010	-0.016	0.014	-0.032	0.022	-0.018
Marg1	Marg2	Marg3	Marg4	Marg5	Marg6	Marg7	Str1	Str2	Str3	
Obj1	0.011	0.057	0.027	0.023	-0.057	0.009	-0.034	0.000	0.043	0.031
Obj2	-0.056	0.046	-0.023	-0.011	-0.001	0.016	0.047	-0.027	0.002	-0.042
Obj3	0.081	0.011	-0.020	0.030	-0.031	-0.093	-0.011	-0.028	-0.004	-0.033
Obj4	-0.046	-0.037	-0.030	0.015	0.019	-0.016	0.029	-0.020	-0.012	0.015
Obj5	0.012	0.010	-0.012	-0.020	0.012	-0.002	-0.019	0.042	-0.026	0.008
Obj6	0.029	-0.044	-0.073	-0.019	0.017	0.057	0.028	-0.030	0.029	-0.004
Obj7	0.077	-0.029	-0.016	-0.023	-0.010	0.052	-0.021	0.057	-0.028	-0.042
Obj8	-0.005	-0.007	-0.031	-0.006	0.044	-0.001	0.014	-0.010	-0.034	-0.001
Obj9	-0.070	-0.021	0.063	-0.046	0.077	-0.013	-0.003	0.005	-0.022	-0.004
Obj10	-0.019	-0.003	0.077	0.021	-0.045	-0.036	-0.004	0.060	-0.001	0.050
Marg1	0.556	-0.028	-0.002	0.052	0.012	-0.006	-0.025	0.024	-0.030	-0.034
Marg2	-0.028	0.682	0.051	-0.083	0.019	0.002	0.047	-0.003	0.008	-0.029
Marg3	-0.002	0.051	0.696	0.071	-0.045	0.016	-0.025	0.002	0.077	0.006
Marg4	0.052	-0.083	0.071	0.682	-0.007	-0.028	-0.008	0.006	0.038	0.009
Marg5	0.012	0.019	-0.045	-0.007	0.709	0.005	-0.022	-0.024	-0.041	0.074

Marg6	-0.006	0.002	0.016	-0.028	0.005	0.692	0.028	-0.042	0.115	0.011
Marg7	-0.025	0.047	-0.025	-0.008	-0.022	0.028	0.766	0.014	-0.081	-0.024
Str1	0.024	-0.003	0.002	0.006	-0.024	-0.042	0.014	0.761	-0.008	-0.005
Str2	-0.030	0.008	0.077	0.038	-0.041	0.115	-0.081	-0.008	0.820	-0.012
Str3	-0.034	-0.029	0.006	0.009	0.074	0.011	-0.024	-0.005	-0.012	0.682
Str4	-0.005	-0.067	-0.033	0.088	0.012	0.036	-0.004	-0.020	0.049	0.076
Str5	0.000	0.031	-0.021	-0.054	0.015	0.007	0.009	0.000	0.032	-0.035
Ang1	0.017	-0.002	-0.027	-0.041	-0.079	-0.051	0.097	0.039	-0.063	0.022
Ang2	-0.077	0.059	-0.012	-0.004	0.000	0.032	0.030	-0.003	0.019	-0.035
Ang3	0.016	0.015	-0.049	-0.015	0.087	-0.063	-0.034	-0.045	-0.066	0.023
	Str4	Str5	Ang1	Ang2	Ang3					
Obj1	-0.072	0.019	-0.065	-0.054	0.047					
Obj2	0.037	0.021	-0.040	0.045	0.027					
Obj3	0.039	0.027	0.002	-0.007	0.023					
Obj4	0.062	-0.048	0.050	-0.018	0.037					
Obj5	-0.036	0.014	0.013	0.025	-0.010					
Obj6	0.038	0.015	-0.034	0.060	-0.016					
Obj7	-0.006	0.018	0.068	-0.065	0.014					
Obj8	-0.024	0.019	0.034	-0.014	-0.032					
Obj9	-0.057	-0.025	0.091	-0.022	0.022					
Obj10	-0.026	-0.061	-0.051	0.031	-0.018					
Marg1	-0.005	0.000	0.017	-0.077	0.016					
Marg2	-0.067	0.031	-0.002	0.059	0.015					
Marg3	-0.033	-0.021	-0.027	-0.012	-0.049					
Marg4	0.088	-0.054	-0.041	-0.004	-0.015					
Marg5	0.012	0.015	-0.079	0.000	0.087					
Marg6	0.036	0.007	-0.051	0.032	-0.063					
Marg7	-0.004	0.009	0.097	0.030	-0.034					
Str1	-0.020	0.000	0.039	-0.003	-0.045					
Str2	0.049	0.032	-0.063	0.019	-0.066					
Str3	0.076	-0.035	0.022	-0.035	0.023					
Str4	0.875	-0.029	-0.007	-0.028	-0.055					
Str5	-0.029	0.831	-0.006	0.032	0.008					
Ang1	-0.007	-0.006	0.774	0.029	0.044					
Ang2	-0.028	0.032	0.029	0.770	0.027					
Ang3	-0.055	0.008	0.044	0.027	0.737					

There are several strategies to evaluate this matrix:

- see how large the residuals are, compared to the original correlations
  - the worst possible model would occur if we extracted no factors and would be the size of the original correlations
  - if the correlations were small to start with, we expect small residuals
  - if the correlations were large to start with, the residuals will be relatively larger (this is not terribly problematic)
- comparing residuals requires squaring them first (because residuals can be both positive and negative)

- the sum of the squared residuals divided by the sum of the squared correlations is an estimate of model fit. Subtracting this from 1.0 means that it ranges from 0 to 1. Values  $> .95$  are an indication of good fit.

Analyzing the residuals means we need to extract only the upper right of the triangle them into an object. We can do this in steps.

```
grmsPAF2_resids <- psych::factor.residuals(GRMSr, grmsPAF2$loadings) #first extract the residuals
grmsPAF2_resids <- as.matrix(grmsPAF2_resids[upper.tri(grmsPAF2_resids)]) #the object has the lower triangle filled with zeros
head(grmsPAF2_resids)
```

```
[,1]
[1,]  0.019934198
[2,] -0.008859929
[3,]  0.055526063
[4,] -0.066056926
[5,] -0.034252440
[6,] -0.023167960
```

One criteria of residual analysis is to see how many residuals there are that are greater than an absolute value of 0.05. The result will be a single column with TRUE if it is  $> |0.05|$  and false if it is smaller. The sum function will tell us how many TRUE responses are in the matrix. Further, we can write script to obtain the proportion of total number of residuals.

```
large.resid <- abs(grmsPAF2_resids) > 0.05
# large.resid
sum(large.resid)
```

```
[1] 55
```

```
round(sum(large.resid)/nrow(grmsPAF2_resids), 3)
```

```
[1] 0.183
```

We learn that there are 55 residuals greater than the absolute value of 0.05. This represents 18% of the total number of residuals.

There are no hard rules about what proportion of residuals can be greater than 0.05. Field recommends that it stay below 50% [Field, 2012].

Another approach to analyzing residuals is to look at their mean. Because of the +/- valences, we need to square them (to eliminate the negative), take the average, then take the square root.

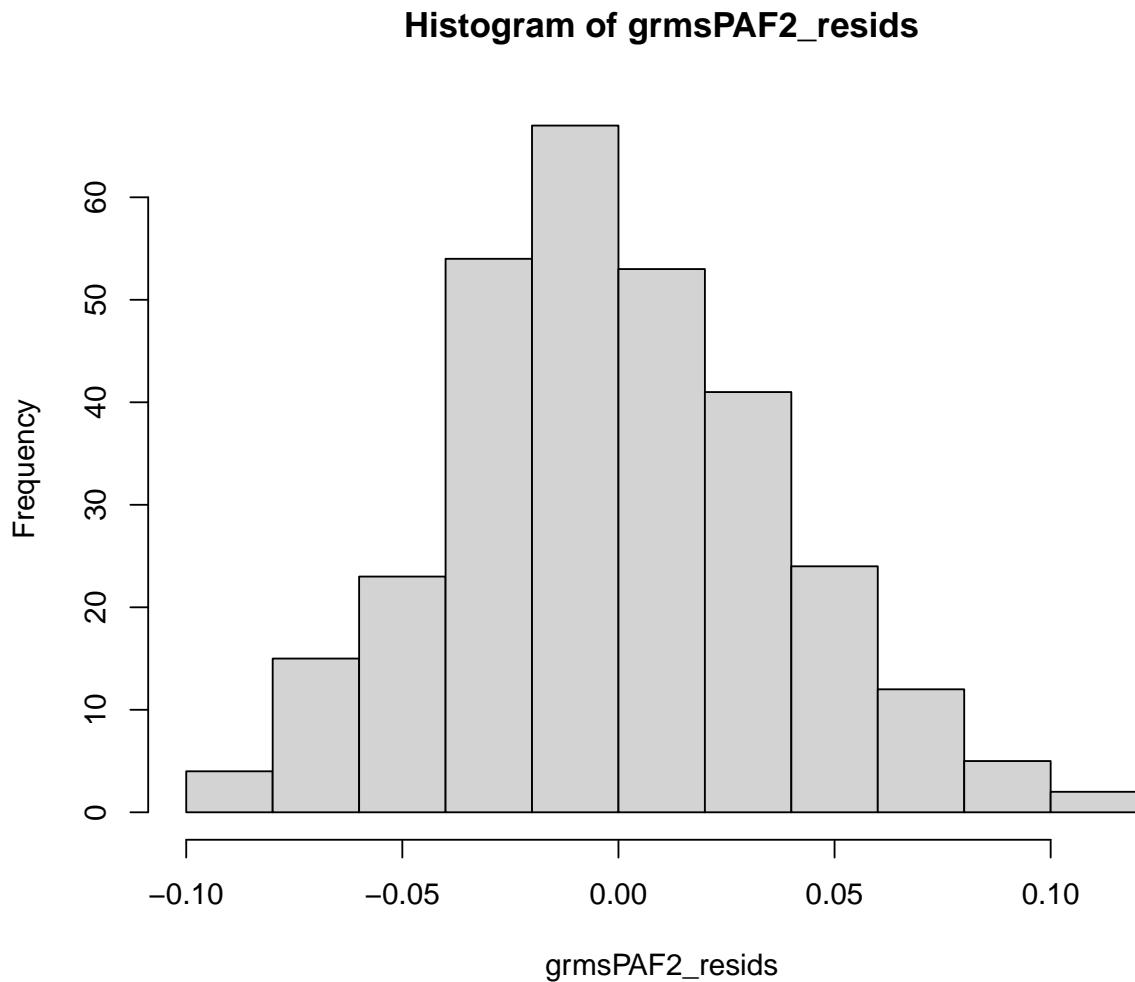
```
round(sqrt(mean(grmsPAF2_resids^2)), 3)
```

```
[1] 0.038
```

While there are no clear guidelines to interpret these, one recommendation is to consider extracting more components if the value is higher than 0.08 [Field, 2012].

Finally, we expect our residuals to be normally distributed. A histogram can help us inspect the distribution.

```
hist(grmsPAF2_resids)
```



Not bad! It looks reasonably normal. No outliers.

#### 9.5.2.2 Quick recap of how to evaluate the # of factors we extracted

- If fewer than 30 variables, the eigenvalue  $> 1$  (Kaiser's) criteria is fine, so long as communalities are all  $> .70$ .
- If sample size  $> 250$  and the average communalities are  $.6$  or greater, this is fine.
- When  $N > 200$ , the scree plot can be used.
- Regarding residuals

- fewer than 50% should have absolute values  $> 0.05$
- model fit should be  $> 0.90$

### 9.5.3 Factor Rotation

The original solution of a principal components or principal axis factor analysis is a set of vectors that best account for the observed covariance or correlation matrix. Each additional component or factor accounts for progressively less and less variance. The solution is efficient (yay) but difficult to interpret (boo).

Thanks to Thurstone's five rules toward a simple structure (circa 1947), interpretation of a matrix is facilitated by *rotation* (multiplying a matrix by a matrix of orthogonal vectors that preserve the communalities of each variable). Both the original matrix and the solution will be orthogonal.

*Parsimony* becomes a statistical consideration (an equation, in fact) and goal and is maximized when each variable has a 1.0 loading on one factor and the rest are zero.

Different rotation strategies emphasize different goals related to parsimony:

*Quartimax* seeks to maximize the notion of variable parsimony (each variable is associated with one factor) and permits the rotation toward a general factor (ignoring smaller factors). *Varimax* maximizes the variance of squared loadings taken over items instead of over factors and *avoids* a general factor.

Rotation improves the interpretation of the factor by maximizing the loading on each variable on one of the extracted factors while minimizing the loading on all other factors. Rotation works by changing the absolute values of the variables while keeping their differential values constant.

There are two big choices (to be made on theoretical grounds):

- Orthogonal rotation if you think that the factors are independent/unrelated.
  - varimax is the most common orthogonal rotation
- Oblique rotation if you think that the factors are related/correlated.
  - oblimin and promax are common oblique rotations

#### 9.5.3.1 Orthogonal rotation

```
# grmsPAF2ORTH <- psych::fa(GRMSr, nfactors = 4, fm = 'pa', rotate =
# 'varimax')
grmsPAF2ORTH <- psych::fa(items, nfactors = 4, fm = "pa", rotate = "varimax")
grmsPAF2ORTH
```

```
Factor Analysis using method = pa
Call: psych::fa(r = items, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
    PA2   PA1   PA3   PA4   h2   u2 com
Obj1  0.47  0.10  0.17  0.13  0.28  0.72  1.5
```

Obj2	0.54	0.14	0.16	-0.11	0.35	0.65	1.4
Obj3	0.38	0.14	0.27	0.09	0.24	0.76	2.2
Obj4	0.47	0.21	0.04	0.10	0.28	0.72	1.5
Obj5	0.40	0.32	0.13	-0.02	0.27	0.73	2.2
Obj6	0.32	0.11	0.01	0.13	0.13	0.87	1.6
Obj7	0.42	0.19	0.21	0.21	0.30	0.70	2.5
Obj8	0.55	0.10	0.00	0.04	0.32	0.68	1.1
Obj9	0.34	0.11	0.07	-0.06	0.14	0.86	1.4
Obj10	0.42	0.04	0.12	0.17	0.22	0.78	1.5
Marg1	0.10	0.61	0.10	0.22	0.44	0.56	1.4
Marg2	0.09	0.49	0.25	0.11	0.32	0.68	1.7
Marg3	0.17	0.50	0.11	0.14	0.30	0.70	1.5
Marg4	0.22	0.47	0.02	0.22	0.32	0.68	1.9
Marg5	0.16	0.50	0.12	0.01	0.29	0.71	1.3
Marg6	0.26	0.42	0.11	0.23	0.31	0.69	2.5
Marg7	0.13	0.47	0.02	0.03	0.23	0.77	1.2
Str1	0.24	0.14	0.40	-0.03	0.24	0.76	1.9
Str2	0.22	0.06	0.28	0.23	0.18	0.82	3.0
Str3	0.05	0.11	0.51	0.20	0.32	0.68	1.4
Str4	0.08	0.09	0.33	0.04	0.13	0.87	1.3
Str5	0.06	0.01	0.40	0.08	0.17	0.83	1.1
Ang1	0.04	0.15	0.31	0.32	0.23	0.77	2.4
Ang2	0.03	0.15	0.10	0.44	0.23	0.77	1.3
Ang3	0.14	0.17	0.16	0.43	0.26	0.74	1.8

	PA2	PA1	PA3	PA4
SS loadings	2.27	2.11	1.22	0.91
Proportion Var	0.09	0.08	0.05	0.04
Cumulative Var	0.09	0.17	0.22	0.26
Proportion Explained	0.35	0.32	0.19	0.14
Cumulative Proportion	0.35	0.67	0.86	1.00

Mean item complexity = 1.7

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

Tucker Lewis Index of factoring reliability = 1.008

RMSEA index = 0 and the 90 % confidence intervals are 0 0.025

BIC = -943.32

Fit based upon off diagonal values = 0.96

## Measures of factor score adequacy

	PA2	PA1	PA3	PA4
Correlation of (regression) scores with factors	0.82	0.81	0.72	0.67
Multiple R square of scores with factors	0.68	0.66	0.52	0.45
Minimum correlation of possible factor scores	0.36	0.33	0.03	-0.10

Essentially, we have the same information as before, except that loadings are calculated after rotation (which adjusts the absolute values of the factor loadings while keeping their differential values constant). Our communality and uniqueness values remain the same. The eigenvalues (SS loadings) should even out, but the proportion of variance explained and cumulative variance (39%) will remain the same.

The *print.psych()* function facilitates interpretation and prioritizes the information about which we care most:

- “cut” will display loadings above .3, this allows us to see
  - if some items load on no factors
  - if some items have cross-loadings (and their relative weights)
- “sort” will reorder the loadings to make it clearer (to the best of its ability...in the case of ties) to which factor/scale it belongs

```
grmsPAF2_table <- psych::print.psych(grmsPAF2ORTH, cut = 0.3, sort = TRUE)
```

```
Factor Analysis using method = pa
Call: psych::fa(r = items, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
    item   PA2   PA1   PA3   PA4   h2   u2 com
Obj8     8 0.55           0.32 0.68 1.1
Obj2     2 0.54           0.35 0.65 1.4
Obj4     4 0.47           0.28 0.72 1.5
Obj1     1 0.47           0.28 0.72 1.5
Obj7     7 0.42           0.30 0.70 2.5
Obj10    10 0.42          0.22 0.78 1.5
Obj5     5 0.40 0.32       0.27 0.73 2.2
Obj3     3 0.38           0.24 0.76 2.2
Obj9     9 0.34           0.14 0.86 1.4
Obj6     6 0.32           0.13 0.87 1.6
Marg1    11     0.61       0.44 0.56 1.4
Marg5    15     0.50       0.29 0.71 1.3
Marg3    13     0.50       0.30 0.70 1.5
Marg2    12     0.49       0.32 0.68 1.7
Marg4    14     0.47       0.32 0.68 1.9
Marg7    17     0.47       0.23 0.77 1.2
Marg6    16     0.42       0.31 0.69 2.5
Str3    20     0.51       0.32 0.68 1.4
Str1    18     0.40       0.24 0.76 1.9
```

Str5	22	0.40	0.17	0.83	1.1
Str4	21	0.33	0.13	0.87	1.3
Str2	19		0.18	0.82	3.0
Ang2	24		0.44	0.23	0.77
Ang3	25		0.43	0.26	0.74
Ang1	23	0.31	0.32	0.23	0.77
					2.4

		PA2	PA1	PA3	PA4
SS loadings		2.27	2.11	1.22	0.91
Proportion Var		0.09	0.08	0.05	0.04
Cumulative Var		0.09	0.17	0.22	0.26
Proportion Explained		0.35	0.32	0.19	0.14
Cumulative Proportion		0.35	0.67	0.86	1.00

Mean item complexity = 1.7

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

Tucker Lewis Index of factoring reliability = 1.008

RMSEA index = 0 and the 90 % confidence intervals are 0 0.025

BIC = -943.32

Fit based upon off diagonal values = 0.96

Measures of factor score adequacy

	PA2	PA1	PA3	PA4
Correlation of (regression) scores with factors	0.82	0.81	0.72	0.67
Multiple R square of scores with factors	0.68	0.66	0.52	0.45
Minimum correlation of possible factor scores	0.36	0.33	0.03	-0.10

In the unrotated solution, most variables loaded on the first component. After rotation, there are four clear components/scales. Further, there is clear (or at least reasonable) component/scale membership for each item and few cross-loadings. Something curious has happened to Str2 – it has no loadings at all! Looking back at the PCA with an orthogonal rotation, Str2 had cross-loadings with two factors.

If this were a new scale and we had not yet established ideas for subscales, the next step is to look back at the items, themselves, and try to name the scales/components. If our scale construction included a priori/planned subscales, here's where we hope the items fall where they were hypothesized to do so. Our simulated data worked perfectly and replicated the four scales that Lewis and Neville [Lewis and Neville, 2015] reported in the article.

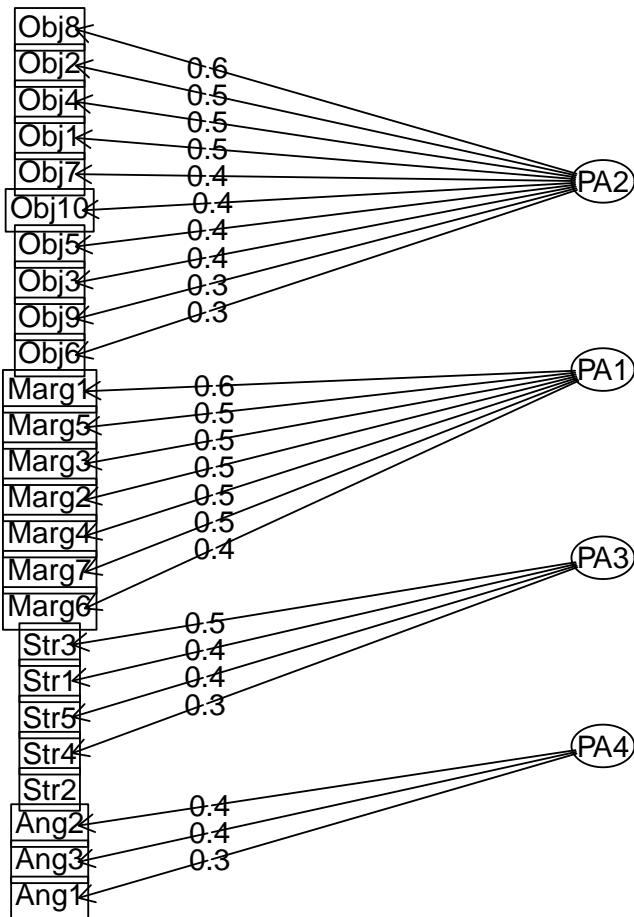
- Assumptions of Beauty and Sexual Objectification

- Silenced and Marginalized
- Strong Woman Stereotype
- Angry Woman Stereotype

We can also create a figure of the result. Note the direction of the arrows from the factor (latent variable) to the items in PAF – in PCA the arrows went from item to component.

```
psych::fa.diagram(grmsPAF2ORTH)
```

## Factor Analysis



We can extract the factor loadings and write them to a table. This can be useful in preparing an APA style table for a manuscript or presentation.

```
# names(grmsPAF2ORTH)
pafORTH_table <- round(grmsPAF2ORTH$loadings, 3)
write.table(pafORTH_table, file = "pafORTH_table.csv", sep = ",",
           col.names = TRUE,
           row.names = FALSE)
pafORTH_table
```

Loadings:

	PA2	PA1	PA3	PA4
Obj1	0.471	0.103	0.166	0.126
Obj2	0.540	0.140	0.162	-0.111
Obj3	0.385	0.136	0.265	
Obj4	0.472	0.206		0.104
Obj5	0.396	0.318	0.130	
Obj6	0.322	0.109		0.133
Obj7	0.421	0.193	0.209	0.210
Obj8	0.551	0.103		
Obj9	0.340	0.112		
Obj10	0.420		0.116	0.167
Marg1	0.105	0.612	0.104	0.218
Marg2		0.486	0.247	0.112
Marg3	0.169	0.495	0.108	0.138
Marg4	0.218	0.471		0.220
Marg5	0.159	0.500	0.124	
Marg6	0.265	0.419	0.107	0.228
Marg7	0.126	0.466		
Str1	0.239	0.140	0.402	
Str2	0.222		0.276	0.225
Str3		0.114	0.512	0.202
Str4			0.331	
Str5			0.398	
Ang1		0.147	0.313	0.324
Ang2		0.150	0.102	0.443
Ang3	0.139	0.175	0.157	0.434

	PA2	PA1	PA3	PA4
SS loadings	2.266	2.105	1.217	0.911
Proportion Var	0.091	0.084	0.049	0.036
Cumulative Var	0.091	0.175	0.224	0.260

### 9.5.3.2 Oblique rotation

Whereas the orthogonal rotation sought to maximize the independence/unrelatedness of the components, an oblique rotation will allow them to be correlated. Researchers often explore both solutions but only report one.

```
# grmsPAF2obl <- psych::fa(GRMSr, nfactors = 4, fm = 'pa', rotate =
# 'oblimin')
grmsPAF2obl <- psych::fa(items, nfactors = 4, fm = "pa", rotate = "oblimin")
```

Loading required namespace: GPArotation

```
grmsPAF2obl
```

Factor Analysis using method = pa  
 Call: psych::fa(r = items, nfactors = 4, rotate = "oblimin", fm = "pa")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	PA2	PA1	PA3	PA4	h2	u2	com
Obj1	0.48	-0.02	0.09	0.07	0.28	0.72	1.1
Obj2	0.55	0.03	0.07	-0.18	0.35	0.65	1.3
Obj3	0.37	0.03	0.21	0.01	0.24	0.76	1.6
Obj4	0.47	0.12	-0.06	0.05	0.28	0.72	1.2
Obj5	0.34	0.27	0.03	-0.12	0.27	0.73	2.2
Obj6	0.33	0.04	-0.06	0.11	0.13	0.87	1.3
Obj7	0.40	0.09	0.13	0.13	0.30	0.70	1.5
Obj8	0.59	-0.01	-0.10	0.01	0.32	0.68	1.1
Obj9	0.34	0.05	0.01	-0.10	0.14	0.86	1.2
Obj10	0.45	-0.08	0.05	0.14	0.22	0.78	1.3
Marg1	-0.05	0.66	0.00	0.08	0.44	0.56	1.0
Marg2	-0.05	0.51	0.18	-0.03	0.32	0.68	1.3
Marg3	0.05	0.52	0.02	0.02	0.30	0.70	1.0
Marg4	0.12	0.48	-0.08	0.12	0.32	0.68	1.3
Marg5	0.03	0.53	0.03	-0.11	0.29	0.71	1.1
Marg6	0.18	0.40	0.01	0.12	0.31	0.69	1.6
Marg7	0.01	0.51	-0.07	-0.07	0.23	0.77	1.1
Str1	0.18	0.06	0.38	-0.14	0.24	0.76	1.8
Str2	0.21	-0.03	0.25	0.16	0.18	0.82	2.7
Str3	-0.03	0.04	0.53	0.09	0.32	0.68	1.1
Str4	0.03	0.04	0.33	-0.04	0.13	0.87	1.1
Str5	0.02	-0.05	0.42	0.00	0.17	0.83	1.0
Ang1	-0.03	0.10	0.31	0.24	0.23	0.77	2.2
Ang2	-0.01	0.13	0.08	0.40	0.23	0.77	1.3
Ang3	0.10	0.12	0.12	0.37	0.26	0.74	1.6

	PA2	PA1	PA3	PA4
SS loadings	2.33	2.30	1.20	0.66
Proportion Var	0.09	0.09	0.05	0.03
Cumulative Var	0.09	0.19	0.23	0.26
Proportion Explained	0.36	0.35	0.19	0.10
Cumulative Proportion	0.36	0.71	0.90	1.00

With factor correlations of

	PA2	PA1	PA3	PA4
PA2	1.00	0.47	0.33	0.09
PA1	0.47	1.00	0.36	0.27
PA3	0.33	0.36	1.00	0.23
PA4	0.09	0.27	0.23	1.00

Mean item complexity = 1.4

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

Tucker Lewis Index of factoring reliability = 1.008

RMSEA index = 0 and the 90 % confidence intervals are 0 0.025

BIC = -943.32

Fit based upon off diagonal values = 0.96

Measures of factor score adequacy

	PA2	PA1	PA3	PA4
Correlation of (regression) scores with factors	0.87	0.88	0.78	0.68
Multiple R square of scores with factors	0.76	0.78	0.61	0.47
Minimum correlation of possible factor scores	0.52	0.56	0.23	-0.07

We can make it a little easier to interpret by removing all factor loadings below .30.

```
psych::print.psych(grmsPAF2obl, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = pa

Call: psych::fa(r = items, nfactors = 4, rotate = "oblimin", fm = "pa")

Standardized loadings (pattern matrix) based upon correlation matrix

item	PA2	PA1	PA3	PA4	h2	u2	com
Obj8	8	0.59			0.32	0.68	1.1
Obj2	2	0.55			0.35	0.65	1.3
Obj1	1	0.48			0.28	0.72	1.1
Obj4	4	0.47			0.28	0.72	1.2
Obj10	10	0.45			0.22	0.78	1.3
Obj7	7	0.40			0.30	0.70	1.5
Obj3	3	0.37			0.24	0.76	1.6
Obj5	5	0.34			0.27	0.73	2.2
Obj9	9	0.34			0.14	0.86	1.2
Obj6	6	0.33			0.13	0.87	1.3
Marg1	11		0.66		0.44	0.56	1.0
Marg5	15		0.53		0.29	0.71	1.1
Marg3	13		0.52		0.30	0.70	1.0
Marg7	17		0.51		0.23	0.77	1.1
Marg2	12		0.51		0.32	0.68	1.3
Marg4	14		0.48		0.32	0.68	1.3
Marg6	16		0.40		0.31	0.69	1.6
Str3	20		0.53		0.32	0.68	1.1

Str5	22	0.42	0.17	0.83	1.0
Str1	18	0.38	0.24	0.76	1.8
Str4	21	0.33	0.13	0.87	1.1
Ang1	23	0.31	0.23	0.77	2.2
Str2	19		0.18	0.82	2.7
Ang2	24		0.40	0.23	0.77
Ang3	25		0.37	0.26	0.74

	PA2	PA1	PA3	PA4
SS loadings	2.33	2.30	1.20	0.66
Proportion Var	0.09	0.09	0.05	0.03
Cumulative Var	0.09	0.19	0.23	0.26
Proportion Explained	0.36	0.35	0.19	0.10
Cumulative Proportion	0.36	0.71	0.90	1.00

With factor correlations of

	PA2	PA1	PA3	PA4
PA2	1.00	0.47	0.33	0.09
PA1	0.47	1.00	0.36	0.27
PA3	0.33	0.36	1.00	0.23
PA4	0.09	0.27	0.23	1.00

Mean item complexity = 1.4

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

Tucker Lewis Index of factoring reliability = 1.008

RMSEA index = 0 and the 90 % confidence intervals are 0 0.025

BIC = -943.32

Fit based upon off diagonal values = 0.96

Measures of factor score adequacy

	PA2	PA1	PA3	PA4
Correlation of (regression) scores with factors	0.87	0.88	0.78	0.68
Multiple R square of scores with factors	0.76	0.78	0.61	0.47
Minimum correlation of possible factor scores	0.52	0.56	0.23	-0.07

The factor structure differs a bit. The Ang items are split between two factors. Again, Str2 has no factor loadings. Additionally, because our specification included “sort=TRUE”, the relative weights wiggled around and so the items are listed in a different order than in the orthogonal rotation.

The oblique rotation allows us to see the correlation between the factors/scales. This was not available in the orthogonal rotation because the assumption of the orthogonal/varimax rotation is that the scales/factors are uncorrelated; hence in the analysis they were fixed to 0.0.

We can see that all the scales have almost no relation with each other. That is, the correlations range between 0.09 to 0.47.

Of course there is always a little complexity. In oblique rotations, there is a distinction between the *pattern* matrix (which reports factor loadings and is comparable to the matrix we interpreted for the orthogonal rotation) and the *structure* matrix (takes into account the relationship between the factors/scales – it is a product of the pattern matrix and the matrix containing the correlation coefficients between the factors/scales). Most interpret the pattern matrix because it is simpler; however it could be that values in the pattern matrix are suppressed because of relations between the factors. Therefore, the structure matrix can be a useful check and some editors will request it.

Obtaining the structure matrix requires two steps. First, multiply the factor loadings with the phi matrix.

```
grmsPAF2obl$loadings %*% grmsPAF2obl$Phi
```

	PA2	PA1	PA3	PA4
Obj1	0.5117768	0.2623780	0.25894758	0.131327445
Obj2	0.5652440	0.2614637	0.22061708	-0.103410346
Obj3	0.4496925	0.2807825	0.34047371	0.098382643
Obj4	0.5118597	0.3325313	0.14601126	0.113266310
Obj5	0.4668753	0.4065387	0.21448069	-0.004545714
Obj6	0.3444817	0.2078771	0.09157767	0.137305996
Obj7	0.4977523	0.3599345	0.32409226	0.222044585
Obj8	0.5538851	0.2351906	0.09469550	0.041663770
Obj9	0.3571910	0.1849101	0.11600401	-0.049653291
Obj10	0.4437917	0.1891586	0.20387746	0.168147070
Marg1	0.2667675	0.6599847	0.24370246	0.254854162
Marg2	0.2443037	0.5404013	0.33965745	0.143501312
Marg3	0.2971065	0.5493661	0.22274243	0.167137489
Marg4	0.3318382	0.5421759	0.16003431	0.245973715
Marg5	0.2812631	0.5274537	0.21175354	0.041944192
Marg6	0.3802024	0.5199470	0.24118171	0.251867664
Marg7	0.2261903	0.4737589	0.10625875	0.052607361
Str1	0.3235913	0.2477221	0.43009276	-0.015627556
Str2	0.2914882	0.2047585	0.34849429	0.231678214
Str3	0.1755369	0.2463368	0.55516637	0.216374544
Str4	0.1555775	0.1626701	0.34786493	0.047592268
Str5	0.1378062	0.1115899	0.40915323	0.087193633
Ang1	0.1466244	0.2675128	0.39209800	0.337320941
Ang2	0.1116268	0.2590113	0.21376799	0.452992854
Ang3	0.2335740	0.3157820	0.28220912	0.444879719

Next, use Field's [2012] function to produce the matrix.

```
# Field's function to produce the structure matrix
factor.structure <- function(fa, cut = 0.2, decimals = 2) {
  structure.matrix <- psych::fa.sort(fa$loadings %*% fa$Phi)
  structure.matrix <- data.frame(ifelse(abs(structure.matrix) < cut,
                                         "", round(structure.matrix, decimals)))
  return(structure.matrix)
}

factor.structure(grmsPAF2obl, cut = 0.3)
```

	PA2	PA1	PA3	PA4
Obj2	0.57			
Obj8	0.55			
Obj4	0.51	0.33		
Obj1	0.51			
Obj7	0.5	0.36	0.32	
Obj5	0.47	0.41		
Obj3	0.45		0.34	
Obj10	0.44			
Obj9	0.36			
Obj6	0.34			
Marg1		0.66		
Marg3		0.55		
Marg4	0.33	0.54		
Marg2		0.54	0.34	
Marg5		0.53		
Marg6	0.38	0.52		
Marg7		0.47		
Str3		0.56		
Str1	0.32	0.43		
Str5		0.41		
Ang1		0.39	0.34	
Str2		0.35		
Str4		0.35		
Ang2			0.45	
Ang3	0.32		0.44	

Although some of the relative values changed, our items were stable regarding their component membership.

#### 9.5.4 Factor Scores

Factor *scores* (PA scores) can be created for each case (row) on each component (column). These can be used to assess the relative standing of one person on the construct/variable to another. We can also use them in regression (in place of means or sums) when groups of predictors correlate so highly that there is multicolliearity.

Computation involves multiplying an individual's item-level response by the component loadings we obtained through the PAF process. The results will be one score per component for each row/case.

```
# in all of this, don't forget to be specifying the dataset that has
# the reverse-coded item replaced
grmsPAF2obl <- psych::fa(items, nfactors = 4, fm = "pa", rotate = "oblimin",
                           scores = TRUE)
head(grmsPAF2obl$scores, 10) #shows us only the first 10 (of N = 2571)
```

	PA2	PA1	PA3	PA4
[1,]	-0.6699528	-0.5575643	-0.55243623	-0.3342807
[2,]	0.2744146	-0.9566521	0.58955654	-0.1364487
[3,]	0.4226985	0.6280590	0.51242085	0.2482677
[4,]	-0.6921922	-1.0455498	-0.01224736	-0.8670751
[5,]	-0.4401667	0.9678210	-1.21941588	-0.1755326
[6,]	-0.1246221	0.8492276	0.11362900	0.5588851
[7,]	0.3611167	0.1384934	-0.67560774	-0.7922718
[8,]	-1.2134910	-0.8242205	0.46592064	1.4000547
[9,]	-0.7439952	-1.1541284	-0.78862308	0.4794409
[10,]	-0.2601972	-0.1055672	-0.61634040	0.6119441

```
items <- cbind(items, grmsPAF2obl$scores) #adds them to our raw dataset
```

To bring this full circle, we can see the correlation of the component scores; the pattern maps onto what we saw previously in the correlations between factors in the oblique rotation.

```
psych::corr.test(items[c("PA1", "PA2", "PA3", "PA4")])
```

```
Call:psych::corr.test(x = items[c("PA1", "PA2", "PA3", "PA4")])
Correlation matrix
  PA1  PA2  PA3  PA4
PA1  1.00  0.60  0.52  0.44
PA2  0.60  1.00  0.49  0.20
PA3  0.52  0.49  1.00  0.44
PA4  0.44  0.20  0.44  1.00
Sample Size
[1] 259
Probability values (Entries above the diagonal are adjusted for multiple tests.)
  PA1  PA2  PA3  PA4
PA1  0    0    0    0
PA2  0    0    0    0
PA3  0    0    0    0
PA4  0    0    0    0
```

To see confidence intervals of the correlations, print with the short=FALSE option

We can extract the factor loadings and write them to a table. This can be useful in preparing an APA style table for a manuscript or presentation.

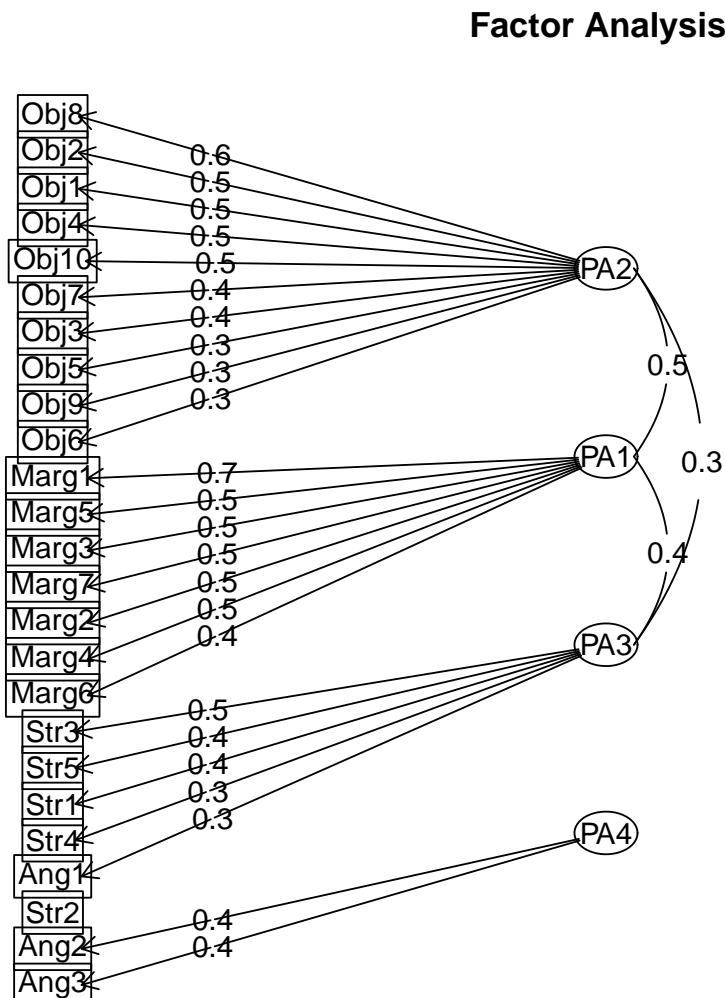
```
# names(grmsPAF2obl)
pafOBL_table <- round(grmsPAF2obl$loadings, 3)
write.table(pafOBL_table, file = "pafOBL_table.csv", sep = ",", col.names = TRUE,
            row.names = FALSE)
pafOBL_table
```

#### Loadings:

	PA2	PA1	PA3	PA4
Obj1	0.484			
Obj2	0.546		-0.179	
Obj3	0.366		0.207	
Obj4	0.472	0.119		
Obj5	0.342	0.265		-0.116
Obj6	0.334		0.108	
Obj7	0.401		0.130	0.130
Obj8	0.590			
Obj9	0.341			
Obj10	0.451		0.135	
Marg1		0.663		
Marg2		0.508	0.179	
Marg3		0.517		
Marg4	0.120	0.482		0.123
Marg5		0.532		-0.113
Marg6	0.177	0.399		0.124
Marg7		0.511		
Str1	0.183		0.379	-0.136
Str2	0.206		0.254	0.162
Str3		0.529		
Str4		0.333		
Str5		0.420		
Ang1		0.102	0.308	0.241
Ang2		0.127		0.401
Ang3	0.101	0.123	0.118	0.375
	PA2	PA1	PA3	PA4
SS loadings	2.090	2.058	1.030	0.579
Proportion Var	0.084	0.082	0.041	0.023
Cumulative Var	0.084	0.166	0.207	0.230

We can also obtain a figure of this PAF with oblique rotation.

```
psych::fa.diagram(grmsPAF2obl)
```



## 9.6 APA Style Results

### Results

The dimensionality of the 25 items from the Gendered Racial Microaggressions Scale for Black Women was analyzed using principal axis factoring. First, data screening evaluated the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00 – values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was .84, indicating acceptable sampling adequacy. The Barlett's Test of Sphericity examines whether the population

correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(300) = 1113.30, p < .001$ , indicating the correlations between items are sufficiently large enough for principal axis factoring. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.01140 and, again, indicated that our data was suitable for the analysis.

Four criteria were used to determine the number of factors to rotate: a priori theory, the scree test, the Eigenvalue-greater-than-one criteria, and the interpretability of the solution. Kaiser's eigenvalue-greater-than-one criteria suggested two components, and, in combination explained 23% of the variance. The scree plot was showed an inflexion that would justified retaining between one and four components. A priori theory based on Lewis and Neville's [2015] psychometric evaluation, suggested four components. Based on the convergence of these decisions, four components were extracted. We investigated each with orthogonal (varimax) and oblique (oblimin) procedures. Given the correspondence of the orthogonal solution with the original research, we selected this as our final model.

The rotated solution, as shown in Table 1 and Figure 1, yielded four interpretable components, each listed with the proportion of variance accounted for: assumptions of beauty and sexual objectification (9%), silenced and marginalized (8%), strong woman stereotype (5%), and angry woman stereotype (4%).

Regarding the Table 1, I would include a table with ALL the values, bolding those with component membership. This will be easy because we exported all those values to a .csv file.

### 9.6.1 Comparing FA and PCA

- FA drives a mathematical solution from which factors are estimated
  - Only FA can estimate underlying factors, but it relies on the various assumptions to be met
- PCA decomposes the original data into a set of linear variates
  - This limits its concern to establishing which linear components exist within the data and how a particular variable might contribute to that component
- Generally, FA and PCA result in similar solutions
  - When there are 30 or more variables and communalities are  $> .7$  for all variables, different solutions are unlikely (Stevens, 2002)
  - When there are  $< 20$  variables and low communalities ( $< .4$ ) different solutions are likely to emerge
  - Both are inferential statistics
- Critics of PCA suggest

- “at best it is a common factor analysis with some error added and at worst an unrecognizable hodgepodge of things from which nothing can be determined” (Cliff, 1987, p. 349)
- PCA should never be described as FA and the resulting components should not be treated as reverently as true, latent variable, *factors*
- To most of us (i.e., scientist-practitioners), the difference is largely from the algorithm used to drive the solutions. This is true for Field [Field, 2012] also, who uses the terms interchangeably. My take: use whichever you like, just be precise in the language describing what you did.

## 9.7 Going Back to the Future: What, then, is Omega?

Now that we've had an introduction to factor analysis, let's revisit the  $\omega$  grouping of reliability estimates. In the context of *psychometrics*, it may be useful to think of factors as scales/subscales where  $g$  refers to the amount of variance in the *general* factor (or total scale score) and subcales to be items that have something in common that is separate from what is  $g$ .

Model-based estimates examine the correlations or covariances of the items and decompose the test variance into that which is:

- common to all items (**g**, a general factor),
- specific to some items (**f**, orthogonal group factors), and
- unique to each item (confounding **s** specific, and **e** error variance)

In the *psych* package

- $\omega_t$  represents the total reliability of the test ( $\omega_t$ )
  - In the *psych* package, this is calculated from a bifactor model where there is one general  $g$  factor (i.e., each item loads on the single general factor), one or more group factors ( $f$ ), and an item-specific factor ( $s$ ).
- $\omega_h$  extracts a higher order factor from the correlation matrix of lower level factors, then applies the Schmid and Leiman (1957) transformation to find the general loadings on the original items. Stated another way, it is a measure of the general factor saturation ( $g$ ; the amount of variance attributable to one comon factor). The subscript “ $h$ ” acknowledges the hierarchical nature of the approach.
  - the  $\omega_h$  approach is exploratory and defined if there are three or more group factors (with only two group factors, the default is to assume they are equally important, hence the factor loadings of those subscales will be equal)
  - Najera Catalan [Najera Catalan, 2019] suggests that  $\omega_h$  is the best measure of reliability when dealing with multiple dimensions.
- $\omega_g$  is an estimate that uses a bifactor solution via the SEM package *lavaan* and tends to be a larger (because it forces all the cross loadings of lower level factors to be 0)
  - the  $\omega_g$  is confirmatory, requiring the specification of which variables load on each group factor

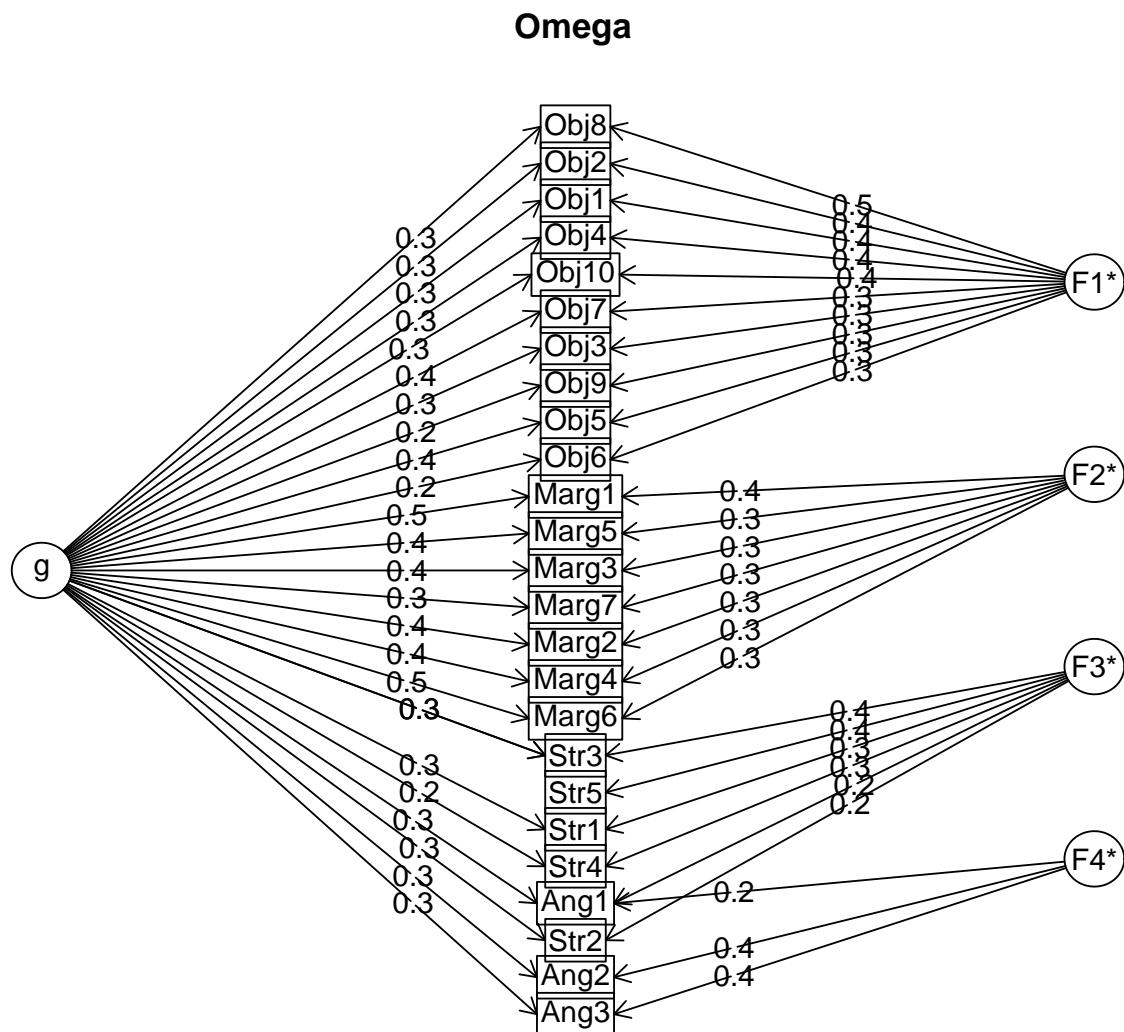
- *psych::omegaSem()* reports both EFA and CFA solutions
  - We will use the *psych::omegaSem()* function

Note that in our specification, we indicate there are two factors. We do not tell it (anywhere!) what items belong to what factors (think, *subscale*). One test will be to see if the items align with their respective factors.

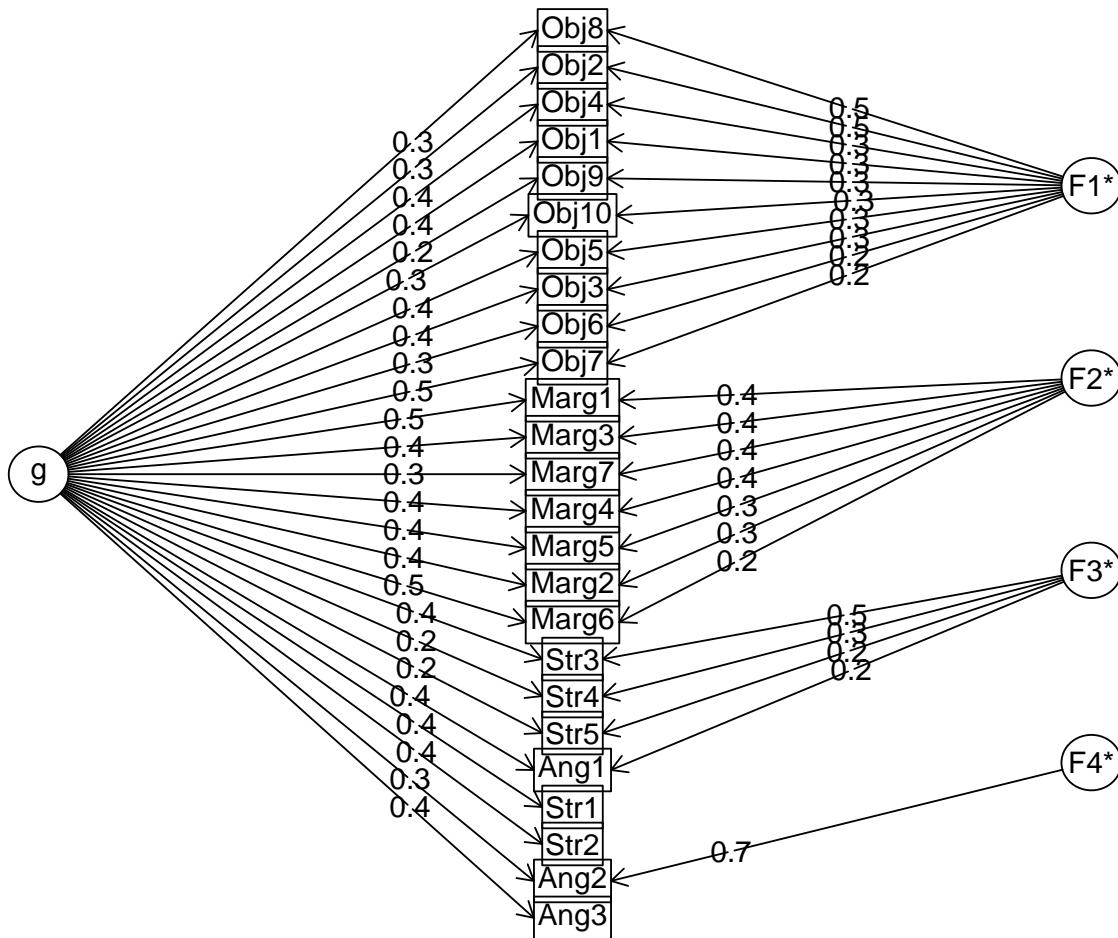
```
# Because we added the component scores to our df (and now it has
# more variables than just our items), I will estimate omegaSem with
# the correlation matrix; I will need to tell it the n.obs

psych::omegaSem(GRMSr, nfactors = 4, n.obs = 259)
```

Warning in lav\_model\_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified.



### Omega from SEM



```

Call: psych::omegaSem(m = GRMSr, nfactors = 4, n.obs = 259)
Omega
Call: omegah(m = m, nfactors = nfactors, fm = fm, key = key, flip = flip,
  digits = digits, title = title, sl = sl, labels = labels,
  plot = plot, n.obs = n.obs, rotate = rotate, Phi = Phi, option = option,
  covar = covar)
Alpha:          0.83
G.6:           0.85
Omega Hierarchical: 0.58
Omega H asymptotic: 0.68
Omega Total      0.85

```

Schmid Leiman Factor loadings greater than 0.2

	<i>g</i>	<i>F1*</i>	<i>F2*</i>	<i>F3*</i>	<i>F4*</i>	<i>h<sup>2</sup></i>	<i>u<sup>2</sup></i>	<i>p<sup>2</sup></i>
<i>Obj1</i>	0.34	0.39				0.27	0.73	0.41
<i>Obj2</i>	0.32	0.45				0.35	0.65	0.29

Obj3	0.35	0.30	0.24	0.76	0.50
Obj4	0.35	0.39	0.28	0.72	0.43
Obj5	0.38	0.28	0.27	0.73	0.53
Obj6	0.23	0.27	0.13	0.87	0.40
Obj7	0.41	0.33	0.30	0.70	0.56
Obj8	0.29	0.48	0.32	0.68	0.26
Obj9	0.21	0.28	0.14	0.86	0.32
Obj10	0.27	0.37	0.22	0.78	0.34
Marg1	0.51	0.42	0.44	0.56	0.58
Marg2	0.45	0.32	0.32	0.68	0.63
Marg3	0.44	0.33	0.31	0.69	0.63
Marg4	0.44	0.31	0.32	0.68	0.61
Marg5	0.41	0.34	0.29	0.71	0.58
Marg6	0.46	0.25	0.31	0.69	0.68
Marg7	0.34	0.32	0.23	0.77	0.51
Str1	0.31	0.33	0.24	0.76	0.40
Str2	0.28	0.22	0.18	0.82	0.46
Str3	0.33	0.45	0.32	0.68	0.33
Str4	0.21	0.29	0.13	0.87	0.35
Str5	0.20	0.35	0.17	0.83	0.23
Ang1	0.31	0.25	0.24	0.23	0.77 0.41
Ang2	0.26		0.39	0.23	0.77 0.29
Ang3	0.34		0.37	0.27	0.73 0.42

With Sums of squares of:

g	F1*	F2*	F3*	F4*
3.00	1.39	0.83	0.73	0.53

general/max 2.17 max/min = 2.6

mean percent general = 0.45 with sd = 0.13 and cv of 0.29

Explained Common Variance of the general factor = 0.46

The degrees of freedom are 206 and the fit is 0.82

The number of observations was 259 with Chi Square = 201.42 with prob < 0.58

The root mean square of the residuals is 0.04

The df corrected root mean square of the residuals is 0.05

RMSEA index = 0 and the 10 % confidence intervals are 0 0.025

BIC = -943.28

Compare this with the adequacy of just a general factor and no group factors

The degrees of freedom for just the general factor are 275 and the fit is 1.69

The number of observations was 259 with Chi Square = 418.67 with prob < 0.00000005

The root mean square of the residuals is 0.08

The df corrected root mean square of the residuals is 0.09

RMSEA index = 0.045 and the 10 % confidence intervals are 0.036 0.053

BIC = -1109.45

## Measures of factor score adequacy

	g	F1*	F2*	F3*	F4*
Correlation of scores with factors	0.78	0.74	0.59	0.65	0.62
Multiple R square of scores with factors	0.61	0.55	0.35	0.42	0.39
Minimum correlation of factor score estimates	0.22	0.10	-0.30	-0.15	-0.23

## Total, General and Subset omega for each subset

	g	F1*	F2*	F3*	F4*
Omega total for total scores and subscales	0.85	0.74	0.75	0.55	0.37
Omega general for total scores and subscales	0.58	0.33	0.47	0.24	0.14
Omega group for total scores and subscales	0.18	0.41	0.27	0.31	0.23

The following analyses were done using the lavaan package

Omega Hierarchical from a confirmatory model using sem = 0.69

Omega Total from a confirmatory model using sem = 0.85

With loadings of

	g	F1*	F2*	F3*	F4*	h2	u2	p2
Obj1	0.41	0.31				0.27	0.73	0.62
Obj2	0.33	0.46				0.32	0.68	0.34
Obj3	0.40	0.26				0.23	0.77	0.70
Obj4	0.40	0.33				0.26	0.74	0.62
Obj5	0.42	0.27				0.24	0.76	0.73
Obj6	0.27	0.22				0.12	0.88	0.61
Obj7	0.51	0.20				0.31	0.69	0.84
Obj8	0.31	0.46				0.31	0.69	0.31
Obj9	0.22	0.29				0.13	0.87	0.37
Obj10	0.34	0.28				0.19	0.81	0.61
Marg1	0.49		0.44			0.43	0.57	0.56
Marg2	0.44		0.30			0.28	0.72	0.69
Marg3	0.43		0.38			0.33	0.67	0.56
Marg4	0.43		0.36			0.32	0.68	0.58
Marg5	0.38		0.34			0.26	0.74	0.56
Marg6	0.51		0.22			0.31	0.69	0.84
Marg7	0.29		0.36			0.21	0.79	0.40
Str1	0.37					0.17	0.83	0.81
Str2	0.38					0.16	0.84	0.90
Str3	0.35		0.53			0.40	0.60	0.31
Str4	0.22		0.34			0.16	0.84	0.30
Str5	0.24		0.23			0.11	0.89	0.52
Ang1	0.36		0.22			0.18	0.82	0.72
Ang2	0.31			0.71	0.59	0.41	0.16	
Ang3	0.42					0.21	0.79	0.84

With sum of squared loadings of:

g	F1*	F2*	F3*	F4*
3.55	1.02	0.85	0.55	0.53

The degrees of freedom of the confirmatory model are 250 and the fit is 256.6602 with p = general/max 3.47 max/min = 1.92  
mean percent general = 0.58 with sd = 0.2 and cv of 0.35  
Explained Common Variance of the general factor = 0.55

Measures of factor score adequacy

	g	F1*	F2*	F3*	F4*
Correlation of scores with factors		0.85	0.72	0.67	0.64 0.77
Multiple R square of scores with factors		0.73	0.51	0.46	0.41 0.59
Minimum correlation of factor score estimates		0.45	0.03	-0.09	-0.18 0.19

Total, General and Subset omega for each subset

	g	F1*	F2*	F3*	F4*
Omega total for total scores and subscales		0.85	0.75	0.75	0.57 0.53
Omega general for total scores and subscales		0.69	0.43	0.45	0.33 0.21
Omega group for total scores and subscales		0.15	0.32	0.30	0.24 0.32

To get the standard sem fit statistics, ask for summary on the fitted object

There's a ton of output! How do we make sense of it?

First, our items aligned perfectly with their respective factors (subscales). That is, it would be problematic if the items switched factors.

Second, we can interpret our results. Like alpha, the omegas range from 0 to 1, where values closer to 1 represent good reliability [Najera Catalan, 2019]. For unidimensional measures,  $\omega_t$  values above 0.80 seem to be an indicator of good reliability. For multidimensional measures with well-defined dimensions we strive for  $\omega_h$  values above 0.65 (and  $\omega_t > 0.8$ ). These recommendations are based on a Monte Carlo study that examined a host of reliability indicators and how their values corresponded with accurate predictions of poverty status. With this in mind, let's examine the output related to our simulated research vignette.

Let's examine the output in the lower portion where the values are "from a confirmatory model using sem."

Omega is a reliability estimate for factor analysis that represents the proportion of variance in the GRMS scale attributable to common variance (rather than error). The omega for the total reliability of the test ( $\omega_t$ ; which included the general factors and the subscale factors) was .85, meaning that 85% of the variance in the total scale is due to the factors and 15% (100% - 85%) is attributable to error.

Omega hierarchical ( $\omega_h$ ) estimates are the proportion of variance in the GRMS score attributable to the general factor, which in effect treats the subscales as error.  $\omega_h$  for the the GRMS total scale was .69 A quick calculation with  $\omega_h$  (.69) and  $\omega_t$  (.85;  $.69/.85 = .81$ ) lets us know that that 81% of the reliable variance in the GRMS total scale is attributable to the general factor.

.69/.85

[1] 0.8117647

Amongst the output is the Cronbach's alpha coefficient (.83). Lewis and Neville [2015] did not report omega results. They reported an alpha of .92 for the version of the GRMS that assessed stress appraisal.

## 9.8 Comparing PFA to Item Analysis and PCA

In the lesson on PCA, we began a table that compared our item analysis (item corrected-total correlations with item-other scale correlations) and PCA results (both orthogonal and oblique). Let's now add our PAF results (both orthogonal and oblique).

In the prior lecture, I saved the file as both .rds and .csv objects. I will bring back in the .rds object and add to it.

```
GRMScomps <- readRDS("GRMS_Comparisons.rds")
grmsPAF2ORTH
```

```
Factor Analysis using method = pa
Call: psych::fa(r = items, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
    PA2   PA1   PA3   PA4   h2   u2 com
Obj1  0.47  0.10  0.17  0.13  0.28  0.72  1.5
Obj2  0.54  0.14  0.16 -0.11  0.35  0.65  1.4
Obj3  0.38  0.14  0.27  0.09  0.24  0.76  2.2
Obj4  0.47  0.21  0.04  0.10  0.28  0.72  1.5
Obj5  0.40  0.32  0.13 -0.02  0.27  0.73  2.2
Obj6  0.32  0.11  0.01  0.13  0.13  0.87  1.6
Obj7  0.42  0.19  0.21  0.21  0.30  0.70  2.5
Obj8  0.55  0.10  0.00  0.04  0.32  0.68  1.1
Obj9  0.34  0.11  0.07 -0.06  0.14  0.86  1.4
Obj10 0.42  0.04  0.12  0.17  0.22  0.78  1.5
Marg1 0.10  0.61  0.10  0.22  0.44  0.56  1.4
Marg2 0.09  0.49  0.25  0.11  0.32  0.68  1.7
Marg3 0.17  0.50  0.11  0.14  0.30  0.70  1.5
Marg4 0.22  0.47  0.02  0.22  0.32  0.68  1.9
Marg5 0.16  0.50  0.12  0.01  0.29  0.71  1.3
Marg6 0.26  0.42  0.11  0.23  0.31  0.69  2.5
Marg7 0.13  0.47  0.02  0.03  0.23  0.77  1.2
Str1  0.24  0.14  0.40 -0.03  0.24  0.76  1.9
Str2  0.22  0.06  0.28  0.23  0.18  0.82  3.0
Str3  0.05  0.11  0.51  0.20  0.32  0.68  1.4
Str4  0.08  0.09  0.33  0.04  0.13  0.87  1.3
Str5  0.06  0.01  0.40  0.08  0.17  0.83  1.1
Ang1  0.04  0.15  0.31  0.32  0.23  0.77  2.4
Ang2  0.03  0.15  0.10  0.44  0.23  0.77  1.3
Ang3  0.14  0.17  0.16  0.43  0.26  0.74  1.8
```

PA2 PA1 PA3 PA4

SS loadings	2.27	2.11	1.22	0.91
Proportion Var	0.09	0.08	0.05	0.04
Cumulative Var	0.09	0.17	0.22	0.26
Proportion Explained	0.35	0.32	0.19	0.14
Cumulative Proportion	0.35	0.67	0.86	1.00

Mean item complexity = 1.7

Test of the hypothesis that 4 factors are sufficient.

df null model = 300 with the objective function = 4.47 with Chi Square = 1113.3  
 df of the model are 206 and the objective function was 0.82

The root mean square of the residuals (RMSR) is 0.04

The df corrected root mean square of the residuals is 0.05

The harmonic n.obs is 259 with the empirical chi square 223.39 with prob < 0.19  
 The total n.obs was 259 with Likelihood Chi Square = 201.39 with prob < 0.58

Tucker Lewis Index of factoring reliability = 1.008

RMSEA index = 0 and the 90 % confidence intervals are 0 0.025

BIC = -943.32

Fit based upon off diagonal values = 0.96

Measures of factor score adequacy

	PA2	PA1	PA3	PA4
Correlation of (regression) scores with factors	0.82	0.81	0.72	0.67
Multiple R square of scores with factors	0.68	0.66	0.52	0.45
Minimum correlation of possible factor scores	0.36	0.33	0.03	-0.10

```
# names(grmsPAF2ORTH) I had to add 'unclass' to the loadings to
# render them into a df
pafORTH_loadings <- data.frame(unclass(grmsPAF2ORTH$loadings))
pafORTH_loadings$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6",
  "Obj7", "Obj8", "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4",
  "Marg5", "Marg6", "Marg7", "Strong1", "Strong2", "Strong3", "Strong4",
  "Strong5", "Angry1", "Angry2", "Angry3") #Item names for joining (and to make sure we know
pafORTH_loadings <- dplyr::rename(pafORTH_loadings, PAF_OR_Mar = PA1, PAF_OR_Obj = PA2,
  PAF_OR_Ang = PA3, PAF_OR_Str = PA4)
# I had to add 'unclass' to the loadings to render them into a df
GRMScomps <- dplyr::full_join(GRMScomps, pafORTH_loadings, by = "Items")

# Now adding the PAF oblique loadings
pafOBLQ_loadings <- data.frame(unclass(grmsPAF2obl$loadings)) #I had to add 'unclass' to the
pafOBLQ_loadings$Items <- c("Obj1", "Obj2", "Obj3", "Obj4", "Obj5", "Obj6",
  "Obj7", "Obj8", "Obj9", "Obj10", "Marg1", "Marg2", "Marg3", "Marg4",
  "Marg5", "Marg6", "Marg7", "Strong1", "Strong2", "Strong3", "Strong4",
  "Strong5", "Angry1", "Angry2", "Angry3")
```

```
# Item names for joining (and to make sure we know which variable is
# which)
pafOBLQ_loadings <- dplyr::rename(pafOBLQ_loadings, PAF_OB_Mar = PA1, PAF_OB_Obj = PA2,
PAF_OB_Ang = PA3, PAF_OB_Str = PA4)

# I had to add 'unclass' to the loadings to render them into a df
GRMScomps <- dplyr::full_join(GRMScomps, pafOBLQ_loadings, by = "Items")

# Writes the table to a .csv file where you can open it with Excel
# and format )
write.csv(GRMScomps, file = "GRMS_Comps.csv", sep = ",", row.names = FALSE,
col.names = TRUE)
```

Warning in write.csv(GRMScomps, file = "GRMS\_Comps.csv", sep = ",", row.names = FALSE, : attempt to set 'col.names' ignored

Warning in write.csv(GRMScomps, file = "GRMS\_Comps.csv", sep = ",", row.names = FALSE, : attempt to set 'sep' ignored

As a research vignette, this has worked extremely well, modeling consistency across the item analysis, principal components analysis (PCA), and principal axis factoring (PAF). That is, items load highest on their own scale (whether it is a component or factor), have no cross-loadings, and do not switch scale memberships from analysis to analysis.

Items	r.drop	Item Corrected-Total ( <b>bold</b> ) and Item-Other:					Principal Components Analysis: Orthogonal Rotations			Principal Components Analysis: Oblique Rotation			Principal Axis Factoring: Orthogonal Rotation			Principal Axis Factoring: Oblique Rotation					
		Objective	Marginalized	Strong	Angry	marginORTH	objORTH	strORTH	angORTH2	marginOBQ	objOBQ	strOBQ	angOBQ	PAF_OR_09	PAF_OR_Mar	PAF_OR_Ang	PAF_OR_Str	PAF_OB_09	PAF_OB_Mar	PAF_OB_Ang	PAF_OB_Str
Obj1	0.41	<b>0.42</b>	0.26	0.26	0.14	<b>0.56</b>	0.07	0.13	0.17	<b>0.57</b>	0.08	0.08	0.13	<b>0.47</b>	0.10	0.17	<b>0.13</b>	<b>0.48</b>	-0.02	0.09	0.07
Obj2	0.40	<b>0.47</b>	0.26	0.22	0.08	<b>0.61</b>	0.14	0.16	-0.15	<b>0.61</b>	0.05	0.10	-0.21	<b>0.54</b>	0.14	0.16	-0.13	<b>0.55</b>	0.03	0.07	-0.18
Obj3	0.42	<b>0.41</b>	0.26	0.29	0.19	<b>0.45</b>	0.12	0.31	0.08	<b>0.43</b>	0.02	0.27	0.02	<b>0.38</b>	0.14	0.27	0.09	<b>0.37</b>	0.03	0.21	0.01
Obj4	0.42	<b>0.45</b>	0.30	0.17	0.19	<b>0.56</b>	0.23	0.09	0.07	<b>0.54</b>	0.16	-0.08	0.02	<b>0.47</b>	0.21	0.04	0.10	<b>0.47</b>	0.12	-0.06	0.05
Obj5	0.43	<b>0.40</b>	0.36	0.21	0.14	<b>0.44</b>	0.37	0.12	-0.08	<b>0.39</b>	0.33	0.05	-0.15	<b>0.40</b>	0.32	0.13	-0.02	<b>0.34</b>	0.27	0.03	-0.12
Obj6	0.28	<b>0.29</b>	0.20	0.13	0.14	<b>0.43</b>	0.08	-0.10	0.23	<b>0.45</b>	0.00	-0.15	0.22	<b>0.32</b>	0.11	0.01	0.13	<b>0.33</b>	0.04	-0.06	0.11
Obj7	0.48	<b>0.43</b>	0.35	0.29	0.26	<b>0.47</b>	0.20	0.22	0.21	<b>0.44</b>	0.10	0.17	0.15	<b>0.42</b>	0.19	0.21	0.21	<b>0.40</b>	0.09	0.13	0.13
Obj8	0.36	<b>0.46</b>	0.23	0.13	0.09	<b>0.64</b>	0.09	-0.04	0.01	<b>0.67</b>	0.00	-0.11	-0.02	<b>0.58</b>	0.10	0.00	0.04	<b>0.59</b>	-0.01	-0.10	0.01
Obj9	0.26	<b>0.31</b>	0.17	0.09	0.09	<b>0.44</b>	0.13	0.06	-0.15	<b>0.43</b>	0.03	0.01	-0.19	<b>0.34</b>	0.11	0.07	-0.06	<b>0.34</b>	0.05	0.01	-0.10
Obj10	0.54	<b>0.30</b>	0.30	0.21	0.15	<b>0.52</b>	0.03	0.08	0.08	<b>0.58</b>	0.22	0.08	-0.13	<b>0.49</b>	0.04	0.12	0.17	<b>0.48</b>	0.05	0.08	0.11
Marg1	0.46	0.31	<b>0.54</b>	0.20	0.27	0.08	<b>0.66</b>	0.09	0.20	<b>0.65</b>	0.09	0.02	0.13	<b>0.61</b>	0.10	0.09	0.22	-0.05	<b>0.66</b>	0.09	0.01
Marg2	0.43	0.27	<b>0.46</b>	0.25	0.28	0.05	<b>0.58</b>	0.23	0.14	-0.07	<b>0.58</b>	0.18	0.06	0.09	<b>0.49</b>	0.25	0.11	-0.05	<b>0.51</b>	0.18	-0.03
Marg3	0.42	0.30	<b>0.49</b>	0.21	0.19	0.15	<b>0.59</b>	0.08	0.12	0.04	<b>0.60</b>	0.01	0.05	0.17	<b>0.50</b>	0.11	0.14	-0.05	<b>0.52</b>	0.02	0.02
Marg4	0.43	0.32	<b>0.46</b>	0.19	0.22	0.20	<b>0.56</b>	-0.01	0.20	0.12	<b>0.55</b>	-0.08	0.15	0.22	<b>0.47</b>	0.02	0.22	0.12	<b>0.48</b>	-0.08	0.12
Marg5	0.40	0.29	<b>0.43</b>	0.20	0.16	0.13	<b>0.62</b>	0.12	-0.06	0.02	<b>0.64</b>	0.05	-0.13	0.18	<b>0.50</b>	0.12	0.01	0.03	<b>0.53</b>	0.03	-0.11
Marg6	0.48	0.37	<b>0.45</b>	0.27	0.24	0.27	<b>0.48</b>	0.06	0.27	0.19	<b>0.45</b>	-0.01	0.21	0.26	<b>0.42</b>	0.11	0.23	0.18	<b>0.40</b>	0.01	0.12
Marg7	0.32	0.24	<b>0.39</b>	0.08	0.18	0.09	<b>0.61</b>	-0.03	-0.03	-0.01	<b>0.65</b>	-0.09	-0.09	0.13	<b>0.47</b>	0.02	0.03	0.01	<b>0.51</b>	-0.07	-0.07
Strong1	0.35	0.32	0.23	<b>0.28</b>	0.14	0.25	0.15	<b>0.55</b>	-0.12	0.19	0.09	<b>0.53</b>	-0.21	0.24	0.14	<b>0.40</b>	-0.03	0.18	0.06	<b>0.38</b>	-0.14
Strong2	0.33	0.27	<b>0.22</b>	0.18	0.26	0.09	<b>0.34</b>	0.34	0.24	-0.10	<b>0.32</b>	0.29	0.22	0.06	<b>0.28</b>	0.23	0.21	-0.03	<b>0.25</b>	0.16	
Strong3	0.24	0.20	0.23	<b>0.36</b>	0.27	0.12	<b>0.63</b>	0.23	-0.03	0.05	<b>0.63</b>	0.15	0.08	0.11	<b>0.51</b>	0.20	-0.02	0.04	<b>0.53</b>	0.01	
Strong4	0.23	0.16	0.21	<b>0.26</b>	0.10	0.04	<b>0.56</b>	0.08	-0.03	0.05	<b>0.57</b>	0.14	0.09	0.09	<b>0.33</b>	0.04	0.05	0.04	<b>0.33</b>	0.04	
Strong5	0.22	0.15	0.11	<b>0.26</b>	0.18	0.04	-0.03	<b>0.57</b>	0.12	0.09	-0.10	<b>0.59</b>	0.06	0.06	0.01	<b>0.40</b>	0.08	0.03	-0.05	<b>0.42</b>	0.01
Angry1	0.32	0.20	0.23	0.29	<b>0.31</b>	0.01	0.12	0.27	<b>0.42</b>	-0.07	0.11	<b>0.37</b>	0.04	0.15	<b>0.31</b>	0.20	-0.03	0.10	<b>0.31</b>	0.24	
Angry2	0.26	0.13	0.24	0.17	<b>0.31</b>	0.01	0.14	0.01	<b>0.70</b>	-0.02	0.08	-0.02	<b>0.70</b>	0.03	0.15	<b>0.10</b>	<b>0.44</b>	0.01	0.13	0.08	<b>0.40</b>
Angry3	0.36	0.26	0.28	0.19	<b>0.34</b>	0.14	0.18	0.09	<b>0.61</b>	0.11	0.10	0.06	<b>0.59</b>	0.14	0.17	<b>0.16</b>	<b>0.43</b>	0.10	0.12	0.02	<b>0.37</b>

Figure 9.3: Comparison of path models for PCA and EFA

## 9.9 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. In the *ReCentering Psych Stats: Psychometrics OER*, it would be ideal if you have selected a dataset you can utilize across the lessons. The least complex is to change the random seed in the research and rework the problem demonstrated in the lesson. The most complex is to use data of your own. In either case, please plan to:

- Properly format and prepare the data.
- Conduct diagnostic tests to determine the suitability of the data for PCA.
- Conducting tests to guide the decisions about number of components to extract.
- Conducting orthogonal and oblique extractions (at least two each with different numbers of components).
- Selecting one solution and preparing an APA style results section (with table and figure).
- Compare your results in light of any other psychometrics lessons where you have used this data (especially the **item analysis** and **PCA** lessons).

### 9.9.1 Problem #1: Play around with this simulation.

Copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results. If PAF is new to you, perhaps you just change the number in “set.seed(240311)” from 240311 to something else. Your results should parallel those obtained in the lecture, making it easier for you to check your work as you go.

### 9.9.2 Problem #2: Conduct a PCA with the Szymanski and Bissonette [2020] research vignette that was used in prior lessons.

The second option involves utilizing one of the simulated datasets available in this OER. Szymanski and Bissonette’s [2020] Perceptions of the LGBTQ College Campus Climate Scale: Development and psychometric evaluation and Keum et al.’s Gendered Racial Microaggressions Scale for Asian American Women [Keum et al., 2018] are ready for PAF analysis. The simulations are available in the chapters in which they are the featured vignette as well as in a simulations appendix at the end of the OER.

### 9.9.3 Problem #3: Try something entirely new.

Using data for which you have permission and access (e.g., IRB approved data you have collected or from your lab; data you simulate from a published article; data from an open science repository; data from other chapters in this OER), complete PAF. The data should allow for at least two (ideally three) components/subscales.

### 9.9.4 Grading Rubric

Using the lecture and workflow (chart) as a guide, please work through all the steps listed in the proposed assignment/grading rubric.

Assignment Component	Points Possible	Points Earned
1. Check and, if needed, format data	5	_____
2. Conduct and interpret the three diagnostic tests to determine if PAF is appropriate as an analysis (KMO, Bartlett’s, determinant).	5	_____
3. Determine how many components to extract (e.g., scree plot, eigenvalues, theory).	5	_____

Assignment Component	Points Possible	Points Earned
4. Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions.	5	_____
5. Conduct an oblique extraction and rotation with a minimum of two different factor extractions.	5	_____
6. Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest.	5	_____
7. APA style results section with table and figure of one of the solutions.	5	_____
8. Explanation to grader	5	_____
<b>Totals</b>	<b>40</b>	_____
<b>40</b>		_____

## 9.10 Homeworked Example

### Screencast Link

For more information about the data used in this homeworked example, please refer to the description and codebook located at the end of the [introduction](#) in first volume of ReCentering Psych Stats.

As a brief review, this data is part of an IRB-approved study, with consent to use in teaching demonstrations and to be made available to the general public via the open science framework. Hence, it is appropriate to share in class. You will notice there are student- and teacher- IDs. These numbers are not connected to the SPU student ID. Rather, the subcontractor who does the course evals for SPU creates a third, not-SPU-related identifier.

This is the same dataset I have been using for many in-class demos. It's great for psychometrics because I actually created a three-factor solution from the institution's course evaluations. We'll get to walk through that process in this class.

Because this is an actual dataset, if you wish to work the problem along with me, you will need to download the data from [LINK TO DATASET](#).

In this homewoRked example I will conduct a principal components analysis. My hope is that the results will support my solution of three dimensions: valued-by-the-student, traditional pedagogy, socially responsive pedagogy.

### 9.10.1 Check and, if needed, format data

```
big <- readRDS("ReC.rds")
```

With the next code I will create an item-level df with only the items used in the three scales.

```
library(tidyverse)
items <- big %>%
  dplyr::select(ValObjectives, IncrUnderstanding, IncrInterest, ClearResponsibilities,
    EffectiveAnswers, Feedback, ClearOrganization, ClearPresentation,
    MultPerspectives, InclusvClassrm, DEIintegration, EquitableEval)
```

Some of the analyses require non-missing data in the df.

```
items <- na.omit(items)
```

Let's check the structure of the data.

```
str(items)
```

```
Classes 'data.table' and 'data.frame': 267 obs. of 12 variables:
 $ ValObjectives      : int 5 5 4 4 5 5 5 4 5 3 ...
 $ IncrUnderstanding   : int 2 3 4 3 4 5 2 4 5 4 ...
 $ IncrInterest        : int 5 3 4 2 4 5 3 2 5 1 ...
 $ ClearResponsibilities: int 5 5 4 4 5 5 4 4 5 3 ...
 $ EffectiveAnswers     : int 5 3 5 3 5 4 3 2 3 3 ...
 $ Feedback             : int 5 3 4 2 5 5 4 4 5 2 ...
 $ ClearOrganization    : int 3 4 3 4 4 5 4 4 5 2 ...
 $ ClearPresentation     : int 4 4 4 2 5 4 4 4 5 2 ...
 $ MultPerspectives      : int 5 5 4 5 5 5 5 5 5 1 ...
 $ InclusvClassrm       : int 5 5 5 5 5 5 5 4 5 3 ...
 $ DEIintegration       : int 5 5 5 5 5 5 5 5 5 2 ...
 $ EquitableEval         : int 5 5 3 5 5 5 5 3 5 3 ...
 - attr(*, ".internal.selfref")=<externalptr>
 - attr(*, "na.action")= 'omit' Named int [1:43] 6 20 106 109 112 113 114 117 122 128 ...
 ..- attr(*, "names")= chr [1:43] "6" "20" "106" "109" ...
```

### 9.10.2 Conduct and interpret the three diagnostic tests to determine if PCA is appropriate as an analysis (KMO, Bartlett's, determinant)

#### 9.10.2.1 KMO

The Kaiser-Meyer-Olkin (KMO) index is an index of *sampling adequacy* to let us know if the sample size is sufficient relative to the statistical characteristics of the data.

General criteria (1974, Kaiser):

- bare minimum of .5
- values between .5 and .7 as mediocre
- values above .9 are superb

```
psych::KMO(items)
```

Kaiser-Meyer-Olkin factor adequacy

Call: psych::KMO(r = items)

Overall MSA = 0.91

MSA for each item =

	ValObjectives	IncrUnderstanding	IncrInterest
	0.94	0.89	0.89
ClearResponsibilities		EffectiveAnswers	Feedback
	0.91	0.93	0.94
ClearOrganization		ClearPresentation	MultPerspectives
	0.94	0.91	0.93
InclusvClassrm		DEIntegration	EquitableEval
	0.86	0.78	0.95

With a KMO of 0.91, the data seems appropriate to continue with the PCA.

### 9.10.2.2 Bartlett's

Barlett's test let's us know if the matrix is an *identity matrix* (i.e., where elements on the off-diagonal would be 0.0 and elements on the diagonal would be 1.0). Stated another way – items only correlate with “themselves” and not other variables.

When  $p < 0.05$  the matrix is not an identity matrix. That is, there are some relationships between variables that can be analyzed.

```
psych::cortest.bartlett(items)
```

R was not square, finding R from data

```
$chisq
[1] 1897.769
```

```
$p.value
[1] 0
```

```
$df
[1] 66
```

The Barlett's test,  $\chi^2(66) = 1897.77, p < 0.001$ , indicating that the correlation matrix is not an identity matrix and, on that dimension, is suitable for analysis.

### 9.10.2.3 Determinant

Multicollinearity or singularity is diagnosed by the determinant. The determinant should be greater than 0.00001. If smaller, then there may be an issue with multicollinearity (variables that are too highly correlated) or singularity (variables that are perfectly correlated).

```
items <- na.omit(items)
det(cor(items))
```

```
[1] 0.0006985496
```

The value of the determinant is 0.0007; greater than 0.00001. We are not concerned with multicollinearity or singularity.

Summary from data screening:

Data screening were conducted to determine the suitability of the data for principal axis factoring. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was 0.91, indicating acceptable sampling adequacy. The Bartlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(66) = 1897.77, p < 0.001$  indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.0007 and, again, indicated that our data was suitable for the analysis.

### 9.10.3 Determine how many components to extract (e.g., scree plot, eigenvalues, theory)

Specify fewer factors than # of items (12). For me it wouldn't run unless it was 6 or fewer.

```
paf1 <- psych::fa(items, nfactors = 6, fm = "pa", max.iter = 100, rotate = "none") # using rat
```

```
maximum iteration exceeded
```

```
paf1
```

```
Factor Analysis using method = pa
Call: psych::fa(r = items, nfactors = 6, rotate = "none", max.iter = 100,
fm = "pa")
```

Standardized loadings (pattern matrix) based upon correlation matrix

	PA1	PA2	PA3	PA4	PA5	PA6	h2	u2	com
ValObjectives	0.52	-0.07	0.16	0.05	0.05	0.10	0.31	0.688	1.3
IncrUnderstanding	0.65	-0.28	0.31	0.03	-0.04	-0.01	0.60	0.400	1.9
IncrInterest	0.74	-0.18	0.50	-0.18	0.13	-0.02	0.87	0.127	2.1
ClearResponsibilities	0.80	-0.09	-0.34	0.10	0.03	-0.06	0.77	0.226	1.4
EffectiveAnswers	0.78	-0.14	-0.13	0.05	0.09	-0.17	0.68	0.315	1.3
Feedback	0.74	0.05	-0.20	-0.12	0.22	0.05	0.65	0.347	1.4
ClearOrganization	0.78	-0.28	-0.13	0.12	-0.07	0.32	0.83	0.175	1.8
ClearPresentation	0.84	-0.21	0.01	0.12	-0.20	-0.11	0.82	0.182	1.3
MultPerspectives	0.81	0.30	-0.12	-0.40	-0.23	0.03	0.97	0.029	2.0
InclusvClassrm	0.64	0.41	0.18	0.21	-0.14	-0.08	0.67	0.327	2.3
DEIIntegration	0.49	0.66	0.13	0.14	0.11	0.10	0.74	0.265	2.2
EquitableEval	0.69	0.09	-0.17	-0.02	0.13	-0.09	0.54	0.462	1.3

	PA1	PA2	PA3	PA4	PA5	PA6
SS loadings	6.11	0.97	0.65	0.31	0.22	0.19
Proportion Var	0.51	0.08	0.05	0.03	0.02	0.02
Cumulative Var	0.51	0.59	0.64	0.67	0.69	0.70
Proportion Explained	0.72	0.11	0.08	0.04	0.03	0.02
Cumulative Proportion	0.72	0.84	0.91	0.95	0.98	1.00

Mean item complexity = 1.7

Test of the hypothesis that 6 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 9 and the objective function was 0.06

The root mean square of the residuals (RMSR) is 0.01

The df corrected root mean square of the residuals is 0.02

The harmonic n.obs is 267 with the empirical chi square 2.97 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 14.81 with prob < 0.096

Tucker Lewis Index of factoring reliability = 0.976

RMSEA index = 0.049 and the 90 % confidence intervals are 0 0.093

BIC = -35.47

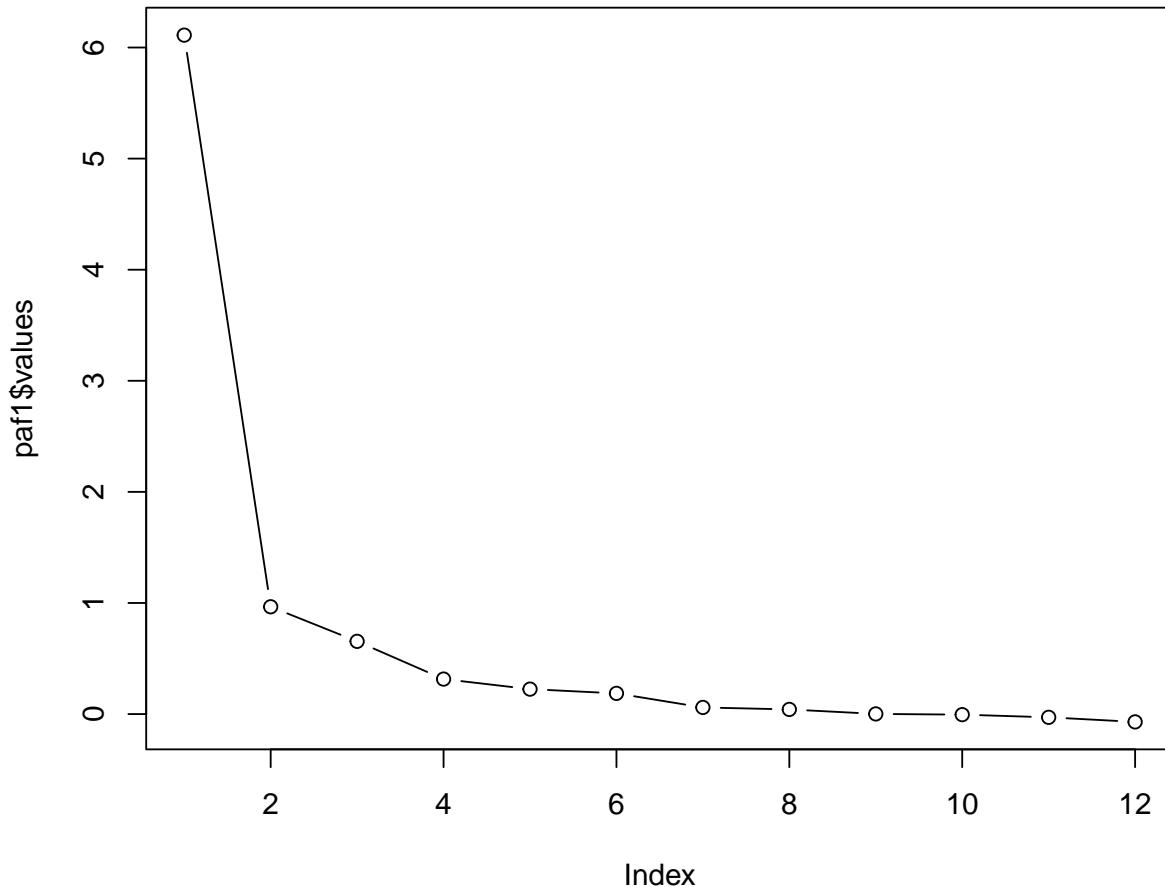
Fit based upon off diagonal values = 1

Measures of factor score adequacy

	PA1	PA2	PA3	PA4	PA5
Correlation of (regression) scores with factors	0.98	0.90	0.88	0.86	0.73
Multiple R square of scores with factors	0.97	0.81	0.77	0.74	0.54
Minimum correlation of possible factor scores	0.93	0.63	0.54	0.48	0.07
				PA6	
Correlation of (regression) scores with factors				0.68	
Multiple R square of scores with factors				0.46	
Minimum correlation of possible factor scores				-0.08	

The eigenvalue-greater-than-one criteria suggests 1 factor (but the second factor has an SSloading of .97).

```
plot(paf1$values, type = "b")
```



The scree plot looks like one factor.

Ugh.

- I want 3 factors (we could think of this as a priori theory); would account for 64% of variance.
- Two could account for 59% of variance.
- Eigenvalues-greater-than-one criteria and scree plot suggests 1 (would account for 51% of variance)

*Note:* The lecture has more on evaluating communalities and uniquenesses and how this information can also inform the number of components we want to extract. Because it is easy to get lost (very lost) I will skip over this for now. If you were to create a measure and use PCA as an exploratory approach to understanding the dimensionality of an instrument, you would likely want to investigate further and report on these.

9.10.4 Conduct an orthogonal extraction and rotation with a minimum of two different factor extractions

## An orthogonal two factor solution

```
pafORTH2f <- psych::fa(items, nfactors = 2, rotate = "varimax")  
pafORTH2f
```

```

Factor Analysis using method = minres
Call: psych::fa(r = items, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix

          MR1   MR2   h2   u2 com
ValObjectives      0.48  0.21  0.27  0.73  1.4
IncrUnderstanding    0.67  0.12  0.46  0.54  1.1
IncrInterest        0.66  0.25  0.50  0.50  1.3
ClearResponsibilities 0.74  0.29  0.63  0.37  1.3
EffectiveAnswers     0.76  0.26  0.64  0.36  1.2
Feedback             0.63  0.38  0.54  0.46  1.7
ClearOrganization     0.79  0.17  0.65  0.35  1.1
ClearPresentation      0.83  0.23  0.75  0.25  1.2
MultPerspectives       0.58  0.55  0.64  0.36  2.0
InclusvClassrm        0.36  0.64  0.54  0.46  1.6
DEIIntegration         0.08  0.86  0.75  0.25  1.0
EquitableEval          0.57  0.40  0.49  0.51  1.8

          MR1   MR2
SS loadings        4.74  2.12
Proportion Var     0.39  0.18
Cumulative Var     0.39  0.57
Proportion Explained 0.69  0.31
Cumulative Proportion 0.69  1.00

```

Mean item complexity = 1.4

Test of the hypothesis that 2 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
df of the model are 43 and the objective function was 0.75

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic n.obs is 267 with the empirical chi square 90.7 with prob < 0.000029  
The total n.obs was 267 with Likelihood Chi Square = 194.26 with prob < 0.0000000000000000

Tucker Lewis Index of factoring reliability = 0.873

RMSEA index = 0.115 and the 90 % confidence intervals are 0.099 0.132

$$BIC = -46$$

Fit based upon off diagonal values = 0.99

Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.94	0.9
Multiple R square of scores with factors	0.89	0.8
Minimum correlation of possible factor scores	0.78	0.6

Sorting the scores into a table can help see the results more clearly. The “cut = #” command will not show the factor scores for factor loading < .30. I would do this “to see”, but I would include all the values in an APA style table.

```
paf_tableORT2f <- psych::print.psych(pafORT2f, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres

Call: psych::fa(r = items, nfactors = 2, rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	h2	u2	com
ClearPresentation	8	0.83		0.75	0.25	1.2
ClearOrganization	7	0.79		0.65	0.35	1.1
EffectiveAnswers	5	0.76		0.64	0.36	1.2
ClearResponsibilities	4	0.74		0.63	0.37	1.3
IncrUnderstanding	2	0.67		0.46	0.54	1.1
IncrInterest	3	0.66		0.50	0.50	1.3
Feedback	6	0.63	0.38	0.54	0.46	1.7
MultPerspectives	9	0.58	0.55	0.64	0.36	2.0
EquitableEval	12	0.57	0.40	0.49	0.51	1.8
ValObjectives	1	0.48		0.27	0.73	1.4
DEIIntegration	11		0.86	0.75	0.25	1.0
InclusvClassrm	10	0.36	0.64	0.54	0.46	1.6

                  MR1    MR2

SS loadings	4.74	2.12
Proportion Var	0.39	0.18
Cumulative Var	0.39	0.57
Proportion Explained	0.69	0.31
Cumulative Proportion	0.69	1.00

Mean item complexity = 1.4

Test of the hypothesis that 2 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 43 and the objective function was 0.75

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic n.obs is 267 with the empirical chi square 90.7 with prob < 0.000029

The total n.obs was 267 with Likelihood Chi Square = 194.26 with prob < 0.0000000000000000

Tucker Lewis Index of factoring reliability = 0.873

RMSEA index = 0.115 and the 90 % confidence intervals are 0.099 0.132

BIC = -46

Fit based upon off diagonal values = 0.99

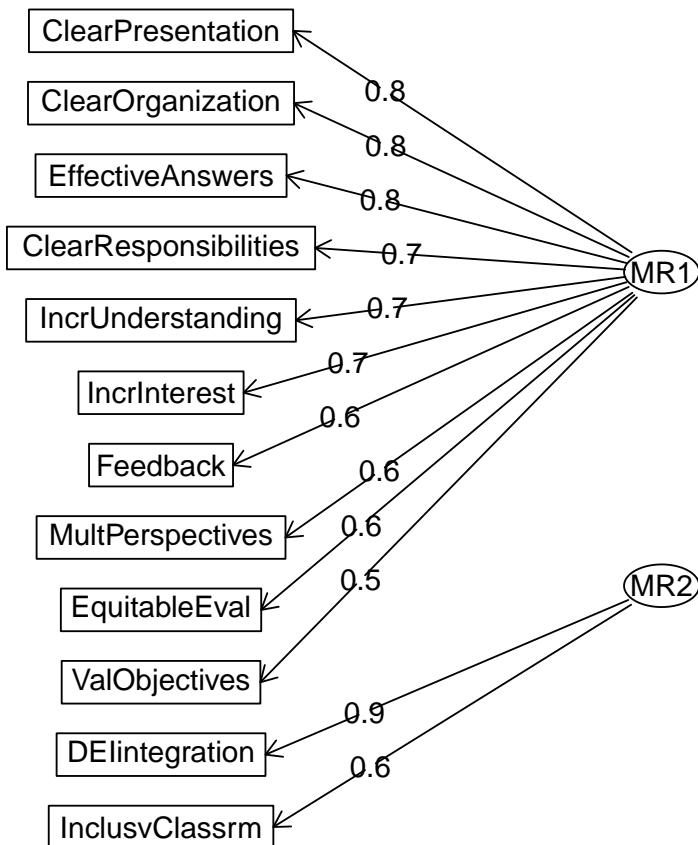
Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.94	0.9
Multiple R square of scores with factors	0.89	0.8
Minimum correlation of possible factor scores	0.78	0.6

F1: Includes everything else. F2: Includes 2 SCR items – DEIintegration, InclsrvClssrm Also: EquitableEval MultPerspectives have high cross-loadings, but end up on the first factor

```
psych::fa.diagram(pafORTH2f)
```

## Factor Analysis



Plotting these figures from the program can facilitate conceptual understanding of what is going

on – and can be a “check” to your work.

In the lecture I made a “biggish deal” about PCA being *components* analysis and PAF being *factor* analysis. Although the two approaches can lead to similar results/conclusions, there are some significant differences “under the hood.” PCA can be thought of more as regression where the items predict the component. Consequently, the arrows went *from* the item, *to* the component.

In PAF, the arrows will go from the factor to the item – because the factors (or latent variables) are assumed to predict the scores on the items (i.e., “depression” would predict how someone rates items that assess hopelessness, sleep, anhedonia, and so forth).

### An orthogonal three factor solution

```
pafORTH3f <- psych::fa(items, nfactors = 3, rotate = "varimax")
pafORTH3f
```

```
Factor Analysis using method = minres
Call: psych::fa(r = items, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
          MR1   MR3   MR2   h2   u2 com
Val0bjectives    0.27  0.43  0.20  0.30  0.70  2.1
IncrUnderstanding 0.27  0.76  0.09  0.66  0.34  1.3
IncrInterest      0.27  0.75  0.25  0.69  0.31  1.5
ClearResponsibilities 0.84  0.24  0.17  0.79  0.21  1.2
EffectiveAnswers   0.67  0.41  0.18  0.64  0.36  1.8
Feedback           0.65  0.26  0.30  0.58  0.42  1.8
ClearOrganization   0.65  0.47  0.10  0.64  0.36  1.9
ClearPresentation   0.62  0.57  0.18  0.74  0.26  2.2
MultPerspectives    0.57  0.29  0.49  0.65  0.35  2.5
InclusvClassrm     0.28  0.30  0.63  0.56  0.44  1.9
DEIIntegration      0.14  0.07  0.85  0.75  0.25  1.1
EquitableEval       0.60  0.23  0.33  0.52  0.48  1.9

          MR1   MR3   MR2
SS loadings    3.37  2.39  1.77
Proportion Var  0.28  0.20  0.15
Cumulative Var  0.28  0.48  0.63
Proportion Explained  0.45  0.32  0.23
Cumulative Proportion 0.45  0.77  1.00
```

Mean item complexity = 1.8

Test of the hypothesis that 3 factors are sufficient.

```
df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77
df of the model are 33 and the objective function was 0.29
```

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.03

The harmonic n.obs is 267 with the empirical chi square 19.55 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 75.2 with prob < 0.0000039

Tucker Lewis Index of factoring reliability = 0.954  
 RMSEA index = 0.069 and the 90 % confidence intervals are 0.049 0.09  
 BIC = -109.18  
 Fit based upon off diagonal values = 1  
 Measures of factor score adequacy

	MR1	MR3	MR2
Correlation of (regression) scores with factors	0.90	0.87	0.89
Multiple R square of scores with factors	0.82	0.76	0.79
Minimum correlation of possible factor scores	0.64	0.52	0.58

```
paf_table0R3f <- psych::print.psych(paf0RTH3f, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres  
 Call: psych::fa(r = items, nfactors = 3, rotate = "varimax")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR3	MR2	h2	u2	com
ClearResponsibilities	4	0.84		0.79	0.21	1.2	
EffectiveAnswers	5	0.67	0.41		0.64	0.36	1.8
Feedback	6	0.65		0.30	0.58	0.42	1.8
ClearOrganization	7	0.65	0.47		0.64	0.36	1.9
ClearPresentation	8	0.62	0.57		0.74	0.26	2.2
EquitableEval	12	0.60		0.33	0.52	0.48	1.9
MultPerspectives	9	0.57		0.49	0.65	0.35	2.5
IncrUnderstanding	2		0.76		0.66	0.34	1.3
IncrInterest	3		0.75		0.69	0.31	1.5
ValObjectives	1		0.43		0.30	0.70	2.1
DEIIntegration	11			0.85	0.75	0.25	1.1
InclusvClassrm	10			0.63	0.56	0.44	1.9

	MR1	MR3	MR2
SS loadings	3.37	2.39	1.77
Proportion Var	0.28	0.20	0.15
Cumulative Var	0.28	0.48	0.63
Proportion Explained	0.45	0.32	0.23
Cumulative Proportion	0.45	0.77	1.00

Mean item complexity = 1.8

Test of the hypothesis that 3 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 33 and the objective function was 0.29

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.03

The harmonic n.obs is 267 with the empirical chi square 19.55 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 75.2 with prob < 0.000039

Tucker Lewis Index of factoring reliability = 0.954

RMSEA index = 0.069 and the 90 % confidence intervals are 0.049 0.09

BIC = -109.18

Fit based upon off diagonal values = 1

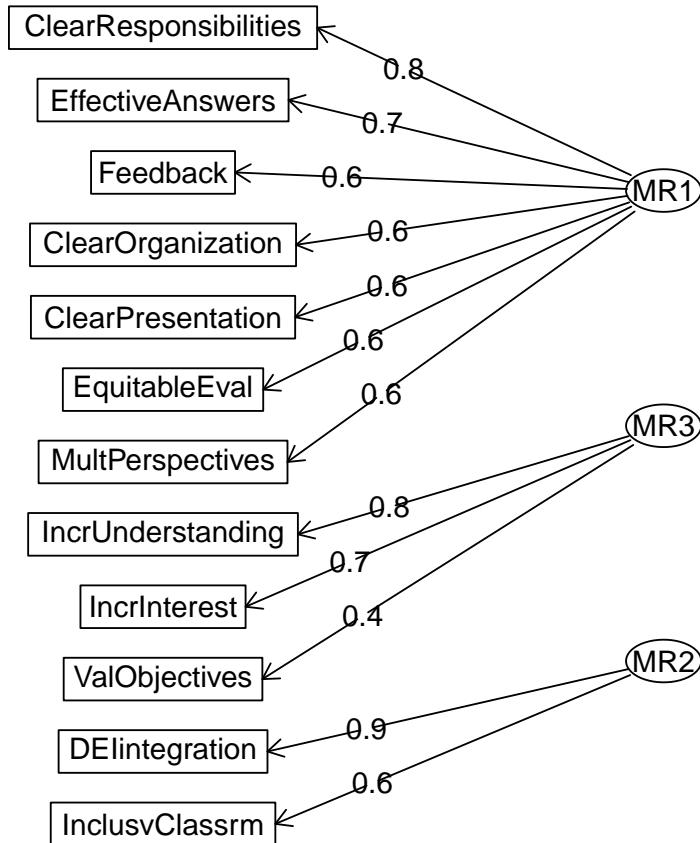
Measures of factor score adequacy

	MR1	MR3	MR2
Correlation of (regression) scores with factors	0.90	0.87	0.89
Multiple R square of scores with factors	0.82	0.76	0.79
Minimum correlation of possible factor scores	0.64	0.52	0.58

F1: Traditional Pedagogy...+MultPerspectives F2: Valued-by-the-Student F3: SCRPed—the 2 items; Note: EquitableEval and MultPerspectivs have some cross-loading with first factor

```
psych::fa.diagram(pafORTH3f)
```

## Factor Analysis



### 9.10.5 Conduct an oblique extraction and rotation with a minimum of two different factor extractions

#### An oblique two factor solution

```
pafOBL2f <- psych::fa(items, nfactors = 2, rotate = "oblimin")
pafOBL2f
```

```
Factor Analysis using method = minres
Call: psych::fa(r = items, nfactors = 2, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
          MR1   MR2   h2   u2 com
ValObjectives      0.50  0.05  0.27  0.73 1.0
IncrUnderstanding    0.73 -0.12  0.46  0.54 1.1
IncrInterest        0.69  0.03  0.50  0.50 1.0
ClearResponsibilities 0.77  0.04  0.63  0.37 1.0
```

EffectiveAnswers	0.80	0.00	0.64	0.36	1.00
Feedback	0.64	0.18	0.54	0.46	1.20
ClearOrganization	0.85	-0.11	0.65	0.35	1.00
ClearPresentation	0.89	-0.05	0.75	0.25	1.00
MultPerspectives	0.55	0.38	0.64	0.36	1.80
InclusvClassrm	0.30	0.55	0.54	0.46	1.50
DEIIntegration	-0.05	0.89	0.75	0.25	1.00
EquitableEval	0.57	0.22	0.49	0.51	1.30

	MR1	MR2
SS loadings	5.32	1.54
Proportion Var	0.44	0.13
Cumulative Var	0.44	0.57
Proportion Explained	0.78	0.22
Cumulative Proportion	0.78	1.00

With factor correlations of

	MR1	MR2
MR1	1.00	0.45
MR2	0.45	1.00

Mean item complexity = 1.2

Test of the hypothesis that 2 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
df of the model are 43 and the objective function was 0.75

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic n.obs is 267 with the empirical chi square 90.7 with prob < 0.000029  
The total n.obs was 267 with Likelihood Chi Square = 194.26 with prob < 0.0000000000000000

Tucker Lewis Index of factoring reliability = 0.873

RMSEA index = 0.115 and the 90 % confidence intervals are 0.099 0.132

$$\text{BIC} = -46$$

Fit based upon off diagonal values = 0.99

### Measures of factor score adequacy

MR1 MR2

Correlation of (regression) scores with factors 0.97 0.91

Multiple R square of scores with factors 0.93 0.83

Minimum correlation of possible factor scores 0.86 0.65

```
paf tableOBL2f <- psych::print.psych(pafOBL2f, cut = 0.3, sort = TRUE)
```

```
Factor Analysis using method = minres  
Call: psych::fa(r = items, nfactors = 2, rotate = "oblimin")
```

Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR2	h2	u2	com
ClearPresentation	8	0.89		0.75	0.25	1.0
ClearOrganization	7	0.85		0.65	0.35	1.0
EffectiveAnswers	5	0.80		0.64	0.36	1.0
ClearResponsibilities	4	0.77		0.63	0.37	1.0
IncrUnderstanding	2	0.73		0.46	0.54	1.1
IncrInterest	3	0.69		0.50	0.50	1.0
Feedback	6	0.64		0.54	0.46	1.2
EquitableEval	12	0.57		0.49	0.51	1.3
MultPerspectives	9	0.55	0.38	0.64	0.36	1.8
ValObjectives	1	0.50		0.27	0.73	1.0
DEIIntegration	11		0.89	0.75	0.25	1.0
InclusvClassrm	10		0.55	0.54	0.46	1.5

	MR1	MR2
SS loadings	5.32	1.54
Proportion Var	0.44	0.13
Cumulative Var	0.44	0.57
Proportion Explained	0.78	0.22
Cumulative Proportion	0.78	1.00

With factor correlations of

	MR1	MR2
MR1	1.00	0.45
MR2	0.45	1.00

Mean item complexity = 1.2

Test of the hypothesis that 2 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
df of the model are 43 and the objective function was 0.75

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic n.obs is 267 with the empirical chi square 90.7 with prob < 0.000029  
The total n.obs was 267 with Likelihood Chi Square = 194.26 with prob < 0.0000000000000000

Tucker Lewis Index of factoring reliability = 0.873

RMSEA index = 0.115 and the 90 % confidence intervals are 0.099 0.132

BIC = -46

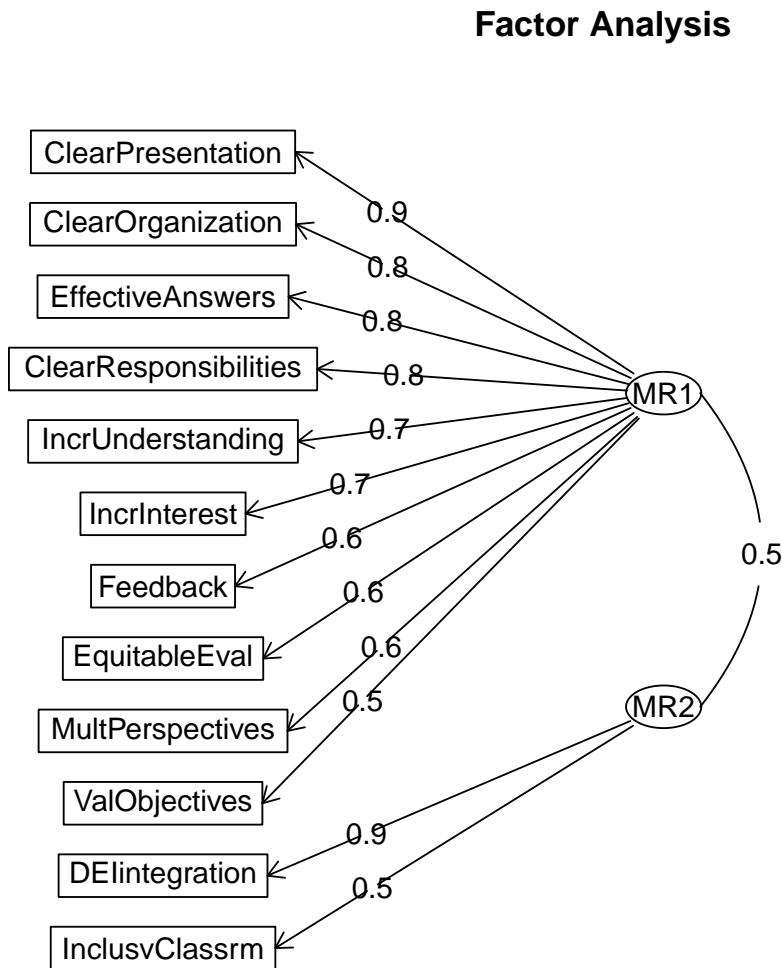
Fit based upon off diagonal values = 0.99

### Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.97	0.91
Multiple R square of scores with factors	0.93	0.83
Minimum correlation of possible factor scores	0.86	0.65

Curiously, there are fewer cross-loadings. F1 has everything except the 2 SCR items which are on F2.

```
psych::fa.diagram(pafOBL2f)
```



With the curved line and value between MR1 and MR2, this figure helps make “allowance” for components to correlate, clear. There was no such path on the orthogonal figures. This is because the rotation required the factors to be uncorrelated.

### An oblique three factor solution

```
pafOBL3f <- psych::fa(items, nfactors = 3, rotate = "oblimin")
pafOBL3f
```

```
Factor Analysis using method = minres
Call: psych::fa(r = items, nfactors = 3, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
          MR1    MR3    MR2    h2    u2  com

```

ValObjectives	0.14	0.40	0.10	0.30	0.70	1.4
IncrUnderstanding	0.03	0.81	-0.05	0.66	0.34	1.0
IncrInterest	0.00	0.78	0.13	0.69	0.31	1.1
ClearResponsibilities	0.97	-0.11	-0.04	0.79	0.21	1.0
EffectiveAnswers	0.68	0.18	-0.01	0.64	0.36	1.1
Feedback	0.69	0.00	0.15	0.58	0.42	1.1
ClearOrganization	0.64	0.27	-0.10	0.64	0.36	1.4
ClearPresentation	0.54	0.41	-0.02	0.74	0.26	1.9
MultPerspectives	0.53	0.07	0.36	0.65	0.35	1.8
InclusvClassrm	0.12	0.21	0.58	0.56	0.44	1.3
DEIIntegration	-0.01	-0.02	0.88	0.75	0.25	1.0
EquitableEval	0.63	-0.01	0.19	0.52	0.48	1.2

	MR1	MR3	MR2
SS loadings	3.80	2.18	1.56
Proportion Var	0.32	0.18	0.13
Cumulative Var	0.32	0.50	0.63
Proportion Explained	0.50	0.29	0.21
Cumulative Proportion	0.50	0.79	1.00

With factor correlations of

MR1	MR3	MR2
MR1	1.00	0.65
MR3	0.65	1.00
MR2	0.43	0.31

Mean item complexity = 1.3

Test of the hypothesis that 3 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 33 and the objective function was 0.29

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.03

The harmonic n.obs is 267 with the empirical chi square 19.55 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 75.2 with prob < 0.000039

Tucker Lewis Index of factoring reliability = 0.954

RMSEA index = 0.069 and the 90 % confidence intervals are 0.049 0.09

BIC = -109.18

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	MR1	MR3	MR2
Correlation of (regression) scores with factors	0.96	0.93	0.91
Multiple R square of scores with factors	0.92	0.86	0.83
Minimum correlation of possible factor scores	0.84	0.71	0.65

```
paf_tableOBL3f <- psych::print.psych(pafOBL3f, cut = 0.3, sort = TRUE)
```

Factor Analysis using method = minres  
 Call: psych::fa(r = items, nfactors = 3, rotate = "oblimin")  
 Standardized loadings (pattern matrix) based upon correlation matrix

	item	MR1	MR3	MR2	h2	u2	com
ClearResponsibilities	4	0.97			0.79	0.21	1.0
Feedback	6	0.69			0.58	0.42	1.1
EffectiveAnswers	5	0.68			0.64	0.36	1.1
ClearOrganization	7	0.64			0.64	0.36	1.4
EquitableEval	12	0.63			0.52	0.48	1.2
ClearPresentation	8	0.54	0.41		0.74	0.26	1.9
MultPerspectives	9	0.53		0.36	0.65	0.35	1.8
IncrUnderstanding	2		0.81		0.66	0.34	1.0
IncrInterest	3		0.78		0.69	0.31	1.1
ValObjectives	1		0.40		0.30	0.70	1.4
DEIIntegration	11			0.88	0.75	0.25	1.0
InclusvClassrm	10			0.58	0.56	0.44	1.3

	MR1	MR3	MR2
SS loadings	3.80	2.18	1.56
Proportion Var	0.32	0.18	0.13
Cumulative Var	0.32	0.50	0.63
Proportion Explained	0.50	0.29	0.21
Cumulative Proportion	0.50	0.79	1.00

With factor correlations of

MR1	MR3	MR2
MR1	1.00	0.65
MR3	0.65	1.00
MR2	0.43	0.31
		1.00

Mean item complexity = 1.3

Test of the hypothesis that 3 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 33 and the objective function was 0.29

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.03

The harmonic n.obs is 267 with the empirical chi square 19.55 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 75.2 with prob < 0.000039

Tucker Lewis Index of factoring reliability = 0.954

RMSEA index = 0.069 and the 90 % confidence intervals are 0.049 0.09

BIC = -109.18

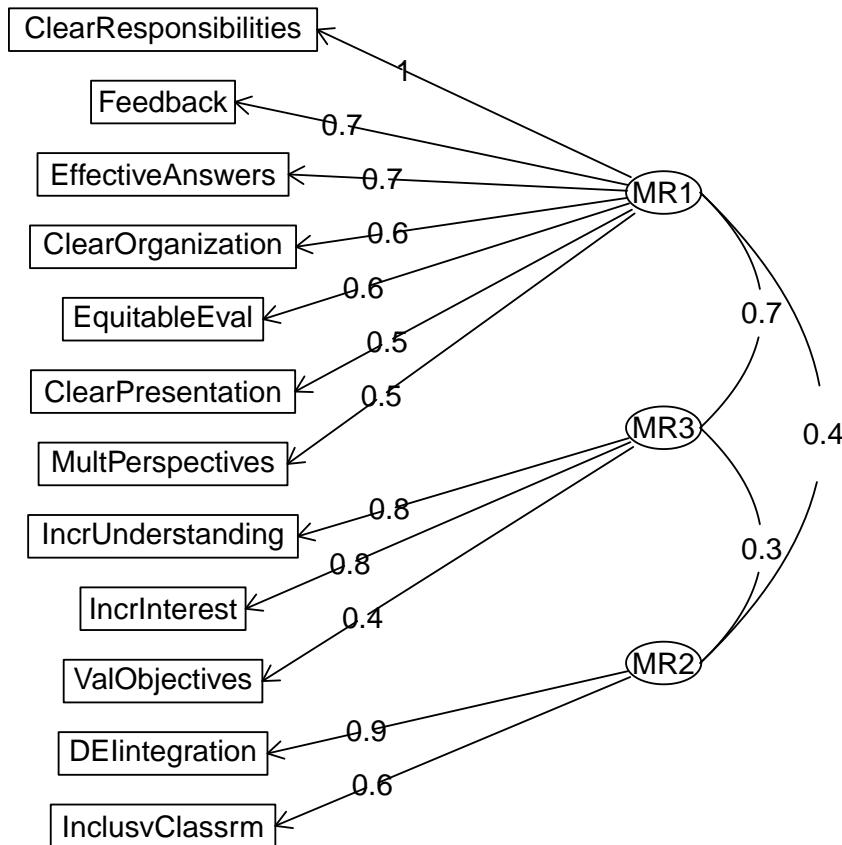
Fit based upon off diagonal values = 1

Measures of factor score adequacy

	MR1	MR3	MR2
Correlation of (regression) scores with factors	0.96	0.93	0.91
Multiple R square of scores with factors	0.92	0.86	0.83
Minimum correlation of possible factor scores	0.84	0.71	0.65

```
psych::fa.diagram(paf0BL3f)
```

## Factor Analysis



Again, pretty similar.

### 9.10.6 Determine which factor solution (e.g., orthogonal or oblique; which number of factors) you will suggest

From the oblique output we see that the correlations between the three subscales range from 0.25 to 0.58. These are high. Therefore, I will choose a 3-component, oblique, solution.

### 9.10.7 APA style results section with table and figure of one of the solutions

The dimensionality of the 12 course evaluation items was analyzed using principal axis factoring (PAF). First, data were screened to determine the suitability of the data for this analyses. Data screening were conducted to determine the suitability of the data for this analyses. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1970) represents the ratio of the squared correlation between variables to the squared partial correlation between variables. KMO ranges from 0.00 to 1.00; values closer to 1.00 indicate that the patterns of correlations are relatively compact and that component analysis should yield distinct and reliable components (Field, 2012). In our dataset, the KMO value was 0.91, indicating acceptable sampling adequacy. The Bartlett's Test of Sphericity examines whether the population correlation matrix resembles an identity matrix (Field, 2012). When the  $p$  value for the Bartlett's test is  $< .05$ , we are fairly certain we have clusters of correlated variables. In our dataset,  $\chi^2(66) = 1897.77, p < 0.001$  indicating the correlations between items are sufficiently large enough for principal components analysis. The determinant of the correlation matrix alerts us to any issues of multicollinearity or singularity and should be larger than 0.00001. Our determinant was 0.0007 and, again, indicated that our data was suitable for the analysis.

Four criteria were used to determine the number of components to extract: a priori theory, the scree test, the eigenvalue-greater-than-one criteria, and the interpretability of the solution. Kaiser's eigenvalue-greater-than-one criteria suggested one component and explained 51% of the variance. The inflexion in the scree plot justified retaining one component. A priorily, we researchers were expecting three components – which would explain 64% of the variance. Correspondingly, we investigated two and three component solutions with orthogonal (varimax) and oblique (oblimin) procedures. Given the significant correlations (ranging from .31 to .65) and the correspondence of items loading on the a priorili hypothesized components, we determined that an oblique, three-component, solution was most appropriate.

The rotated solution, as shown in Table 1 and Figure 1, yielded thre interpretable components, each listed with the proportion of variance accounted for: traditional pedagogy (32%), valued-by-me (18%), and socially and culturally responsive pedagogy (13%).

Regarding the Table 1, I would include a table with ALL the values, bolding those with component membership. This is easy, though, because we can export it to a .csv file and

```
pafOBL3fb <- psych::fa(items, nfactors = 3, rotate = "oblimin")
paf_tableOBL3fb <- psych::print.psych(pafOBL3fb, sort = TRUE)
```

Factor Analysis using method = minres

```

Call: psych::fa(r = items, nfactors = 3, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
    item   MR1   MR3   MR2   h2   u2 com
ClearResponsibilities  4  0.97 -0.11 -0.04  0.79  0.21 1.0
Feedback                 6  0.69  0.00  0.15  0.58  0.42 1.1
EffectiveAnswers        5  0.68  0.18 -0.01  0.64  0.36 1.1
ClearOrganization       7  0.64  0.27 -0.10  0.64  0.36 1.4
EquitableEval           12 0.63 -0.01  0.19  0.52  0.48 1.2
ClearPresentation        8  0.54  0.41 -0.02  0.74  0.26 1.9
MultPerspectives         9  0.53  0.07  0.36  0.65  0.35 1.8
IncrUnderstanding        2  0.03  0.81 -0.05  0.66  0.34 1.0
IncrInterest             3  0.00  0.78  0.13  0.69  0.31 1.1
ValObjectives            1  0.14  0.40  0.10  0.30  0.70 1.4
DEIIntegration           11 -0.01 -0.02  0.88  0.75  0.25 1.0
InclusvClassrm          10  0.12  0.21  0.58  0.56  0.44 1.3

```

	MR1	MR3	MR2
SS loadings	3.80	2.18	1.56
Proportion Var	0.32	0.18	0.13
Cumulative Var	0.32	0.50	0.63
Proportion Explained	0.50	0.29	0.21
Cumulative Proportion	0.50	0.79	1.00

With factor correlations of

MR1	MR3	MR2
MR1 1.00	0.65	0.43
MR3 0.65	1.00	0.31
MR2 0.43	0.31	1.00

Mean item complexity = 1.3

Test of the hypothesis that 3 factors are sufficient.

df null model = 66 with the objective function = 7.27 with Chi Square = 1897.77  
 df of the model are 33 and the objective function was 0.29

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.03

The harmonic n.obs is 267 with the empirical chi square 19.55 with prob < 0.97  
 The total n.obs was 267 with Likelihood Chi Square = 75.2 with prob < 0.000039

Tucker Lewis Index of factoring reliability = 0.954

RMSEA index = 0.069 and the 90 % confidence intervals are 0.049 0.09

BIC = -109.18

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	MR1	MR3	MR2
Correlation of (regression) scores with factors	0.96	0.93	0.91

Multiple R square of scores with factors	0.92	0.86	0.83
Minimum correlation of possible factor scores	0.84	0.71	0.65

```
pafOBL3fb_table <- round(pafOBL3fb$loadings, 3)
write.table(pafOBL3fb_table, file = "pafOBL3f_table.csv", sep = ",",
            row.names = FALSE)
pafOBL3fb_table
```

Loadings:

	MR1	MR3	MR2
ValObjectives	0.140	0.398	0.103
IncrUnderstanding		0.810	
IncrInterest		0.784	0.128
ClearResponsibilities	0.971	-0.108	
EffectiveAnswers	0.676	0.182	
Feedback	0.686		0.146
ClearOrganization	0.640	0.267	
ClearPresentation	0.543	0.405	
MultPerspectives	0.527		0.363
InclusvClassrm	0.121	0.207	0.580
DEIIntegration			0.880
EquitableEval	0.629		0.185

	MR1	MR3	MR2
SS loadings	3.283	1.758	1.339
Proportion Var	0.274	0.146	0.112
Cumulative Var	0.274	0.420	0.532

### 9.10.8 Explanation to grader

# **Confirmatory Factor Analysis**



# Chapter 10

## CFA: First Order Models

### [Screencasted Lecture Link](#)

This is the first in our series on confirmatory factor analysis (CFA).

Our goal is:

- Comparison of CFA to EFA/PCA
- Identify issues in specifying models
- Specifying and running first order models:
  - unidimensional
  - multidimensional
- Interpreting output
- Comparing two versions (unidimensional, multidimensional) of a first-order model

### 10.1 Navigating this Lesson

This lesson is just over two hours. I would add another two hours to work through and digest the materials.

While the majority of R objects and data you will need are created within the R script that sources the chapter, occasionally there are some that cannot be created from within the R framework. Additionally, sometimes links fail. All original materials are provided at the [Github site](#) that hosts the book. More detailed guidelines for ways to access all these materials are provided in the OER's [introduction](#)

#### 10.1.1 Learning Objectives

Focusing on this week's materials, make sure you can:

- Compare and contrast EFA and CFA
- Identify the components of item-level variance in CFA
- Specify CFA measurement models

- Interpret fit indices (e.g., Chi-square, CFI, RMSEA)
- Interpret statistics used to compare two CFA models

### 10.1.2 Planning for Practice

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The least complex is to change the random seed and rework the problem demonstrated in the lesson. The results *should* map onto the ones obtained in the lecture.

The second option comes from the “the back of the book” where a **chapter** contains simulated data for all of the examples worked in this volume. Any of these is available for CFA.

As a third option, you are welcome to use data to which you have access and is suitable for CFA. These could include other simulated data, data available through open science repositories, or your own data (presuming you have permission to use it).

The suggestion for practice spans this chapter and the **next**. From this assignment, you should plan to:

- Prepare the data frame for CFA.
- Specify and run unidimensional and single order (with correlated factors) models.
  - In the next chapter, you will add the specification, evaluation, and write-up of second-order and bifactor models.
- Narrate the adequacy of fit with  $\chi^2$ , CFI, RMSEA, SRMR
  - Write a mini-results section for each
- Compare model fit with  $\chi^2\Delta$ , AIC, and BIC.
- Write an APA style results sections with table(s) and figures.

### 10.1.3 Readings & Resources

In preparing this chapter, I drew heavily from the following resource(s). Other resources are cited (when possible, linked) in the text with complete citations in the reference list.

Byrne, B. M. (2016). Structural equation modeling with AMOS: Basic concepts, applications, and programming (3rd ed.). Routledge. <http://ebookcentral.proquest.com/lib/spu/detail.action?docID=4556523>

- Chapter 1, Structural Equation Modeling: The basics
- Chapter 3, Application 1: Testing the Factorial Validity of a Theoretical Construct (First-Order CFA Model)
- Chapter 4, Application 2: Testing the Factorial Validity of a Measurement Scale (First-Order CFA Model)

Dekay, Nicole (2021). Quick Reference Guide: The statistics for psychometrics <https://www.humananalysts.com/quick-reference-guide-the-statistics-for-psychometrics>

Kline, R. (2016). Principles and practice of structural equation modeling (Fourth ed., Methodology in the social sciences). New York: The Guilford Press.

- Chapter 9: Specification and Identification of Confirmatory Factor Analysis Models
- Chapter 13: Analysis of Confirmatory Factor Analysis Models
- Chapter 12: Global Fit Testing

Rosseel, Y. (2019). The *lavaan* tutorial. Belgium: Department of Data Analysis, Ghent University. <http://lavaan.ugent.be/tutorial/tutorial.pdf>

- “The model syntax” pp. 3 - 4
- “A first example: confirmatory factor analysis (CFA)” pp. 4-8.

#### 10.1.4 Packages

The packages used in this lesson are embedded in this code. When the hashtags are removed, the script below will (a) check to see if the following packages are installed on your computer and, if not (b) install them.

```
#will install the package if not already installed
#if(!require(lavaan)){install.packages("lavaan")}
#if(!require(lavaanPlot)){install.packages("lavaanPlot")}
#if(!require(psych)){install.packages("psych")}
#if(!require(semTable)){install.packages("semTable")}
```

## 10.2 Two Broad Categories of Factor Analysis: Exploratory and Confirmatory

Kline [2016] described confirmatory factor analysis as “exactly half that of SEM – the other half comes from regression analysis” (p. 189).

### 10.2.1 Common to Both Exploratory and Confirmatory Approaches

In both exploratory and confirmatory approaches, the variance of each indicator/item is divided into **common** and **unique** variance. When we assume that variance is 1.0, the common variance becomes the communality. If we have 8 items, we will have 8 communalities and this represents the common variance explained by the factors or components.

- **Common variance** is shared among the indicators and serves as a basis for observed covariances among them that depart, meaningfully, from zero. We generally assume that
  - common variance is due to the factors, and
  - there will be fewer factors than the number of indicators/items (after all, there is no point in retaining as many factors [explanatory entities] as there are entities to be explained [indicators/items])

- the proportion of total variance that is shared is the **communality** (estimated by  $h^2$ ); if  $h^2 = .70$ , then 70% of the total indicator variance is common and potentially explained by the factors
- **Unique variance** consists of
  - **specific variance**: systematic variance that is not explained by any factor in the model
  - **random measurement error**
  - **method variance** is not pictured, but could be another source of unique variance
- In factor analysis, summing the communalities represents the total common variance (a portion of the total variance), but not the total variance.
  - Factor analysis, then, aligns well with classic test theory and classic approaches to understanding reliability (observed score = true score + error).
  - The inclusion of error is illustrated well in the classic illustrations of CFA and SEM where each item/indicator includes common variance (from the factor) and error variance.

Recall that principal components analysis (PCA is not factor analysis) one of the key distinctions is that all variance is common variance (there is no unique variance). Total common variance is equal to the total variance explained, which in turn is equal to the total variance.

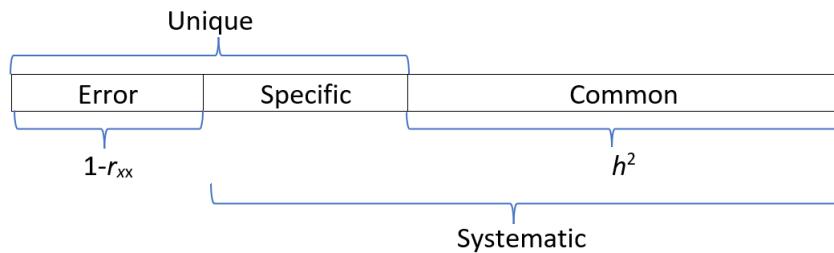


Figure 10.1: Figure illustrating the unique and common variance associated with a factor

### 10.2.2 Differences between EFA and CFA

- **A priori specification of the number of factors**
  - EFA requires no a priori specification; prior to extraction an EFA program will extract as many factors as indicators. Typically, in subsequent analyses, the researchers specifies how many factors to extract.
  - CFA requires researchers to specify the exact number of factors.
- **The degree of “exact correspondence” between indicators/items and factors/scales**
  - EFA is an **unrestricted measurement model** That is, indicators/items depend on (theoretically, measure) all factors. The direct effects from factors to indicators are *pattern coefficients*. Kline [2016] says that most refer to these as *factor loadings* or just *loadings* but because he believes these terms are ambiguous, he refers to the direct effects as *pattern coefficients*. We assign them to factors based on their highest loadings (and

hopefully no cross-loadings). Depending on whether we select an orthogonal or oblique relationship, correlations between factors will be permitted or suppressed.

- CFA is a **restricted measurement model**. The researcher specifies the factor(s) on which each indicator/item(s) depends (recall, the causal direction in CFA is from factor to indicators/items.)
- **Identification status** The *identification* of a model has to do with whether it is theoretically possible for a computer to derive a unique set of model parameter estimates. Identification is related to model *degrees of freedom*; we will later explore under-, just-, and over-identified models. For now:
  - EFA models with multiple factors are *unidentified* because they will have more free parameters than observations. Thus, there is no unique set of statistical estimates for the multifactor EFA model, consequently this requires the rotation phase in EFA.
  - CFA models must be identified before they can be analyzed so there is only one unique set of parameter estimates. Correspondingly, there is no rotation phase in CFA.
- **Sharing variances**
  - In EFA the specific variance of each indicator is not shared with that of any other indicator.
  - In CFA, the researchers can specify if variance is shared between certain pairs of indicators (i.e., error covariances).

### 10.2.3 On the relationship between EFA and CFA

Kline [2016] admonishes us to not overinterpret the labels “exploratory” and “confirmatory”. Why?

- EFA requires no a priori hypotheses about the relationship between indicators/items and factors, but researchers often expect to specify a predetermined number of factors.
- CFA is not strictly confirmatory. After initial runs, many researchers modify models and hypotheses.

CFA is not a verification or confirmation of EFA results for the same data and number of factors. Kline [2016] does not recommend that researchers follow a model retained from EFA. Why?

- It is possible that the CFA model will be rejected. Oftentimes this is because the secondary coefficients (i.e., non-primary pattern coefficients) accounted for a significant proportion of variance in the model. When they are constrained to 0.0 in the CFA model, the model fit will suffer.
- If the CFA model is retained, then it is possible that both EFA and CFA capitalized on chance variation. Thus, if verification via CFA is desired, it should be evaluated through a replication sample.

## 10.3 Exploring a Standard CFA Model

The research vignette for today is a fairly standard CFA model.

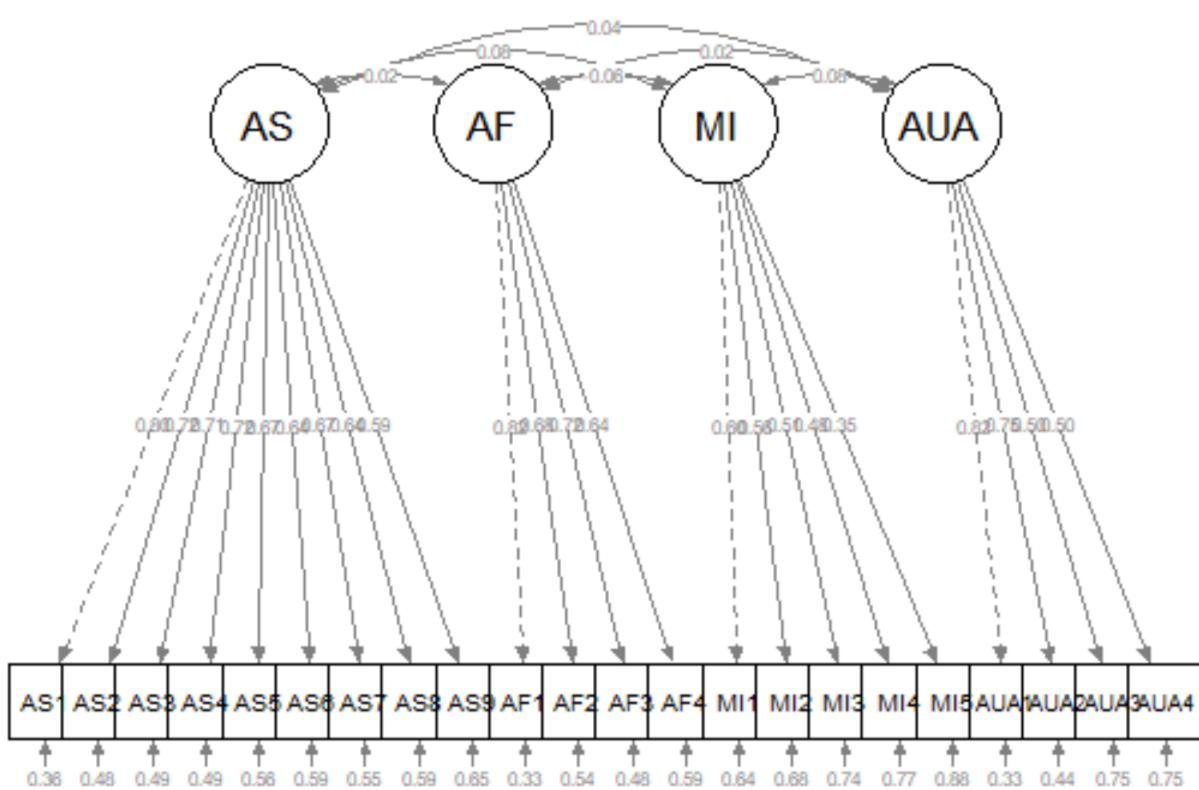


Figure 10.2: Image of the GRMSAAW represented as a standard CFA model

The image represents represents the hypothesis that  $AS_1 - AS_9$ ,  $AF_1 - AF_4$ ,  $MI_1 - MI_5$ , and  $AUA_1 - AUA_4$  measure, respectively, the AS, AF, MI, and AUA factors, which are assumed to covary. Specifically,in this model:

1. Each indicator is continuous with two causes:  $AS \rightarrow AS_1 \leftarrow E_1$ 
  - a single factor that the indicator is supposed to measure, and
  - all unique sources of influence represented by the error term
2. The error terms are independent of each other and of the factors
3. All associations are linear and the factors covary.
  - hence, the symbol for an unanalyzed association is a solid line (upgraded from the dashed one in the EFA)
4. Each item has a single *pattern coefficient* (i.e., often more casually termed as a “factor loading”)
  - All other potential pattern coefficients are set to “0.00.” These are *hard hypotheses* and are specified by their absence (i.e., not specified in the code or in the diagram).
5. *Structure coefficients* are the Pearson correlations between factors and continuous indicators. They reflect any source of association, causal or non causal. Sometimes the association is an undirected, back-door path. There is no pattern coefficient for  $AS_2 \leftrightarrow AF$ . BUT, there is a connection from  $AS_2$  to  $AF$  via the  $AS \leftrightarrow AF$  covariance.
6. *Scaling constants* (aka *unit loading identification [ULI] constraints*) are necessary to scale the factors in a metric related to that of the explained (common) variance of the corresponding indicator, or *reference (marker) variable*. In the figure these are the dashed-line paths from  $AS \rightarrow AS_1$ ,  $AF \rightarrow AF_1$ ,  $MI \rightarrow MI_1$  and  $AUA \rightarrow AUA_1$ .
  - Selecting the reference marker variable is usually arbitrary and selected by the computer program as the first (or last) variable in the code/path. So long as all the indicator variables of the same factor have equally reliable scores, this works satisfactorily.
  - Additional scaling constants are found for each of the errors and indicators.

### 10.3.1 Model Identification for CFA

SEM, in general, requires that all models be *identified*. Measurement models analyzed in CFA share this requirement, but identification is more straightforward than in other models.

Standard CFA models are sufficiently identified when:

1. A single factor model has at least three indicators, or
2. In a model with two or more factors, each factor has two or more indicators.
  - Note: It is better to have at least three to five indicators per factor to prevent technical problems with statistical identification.

Identification becomes much more complicated than this, but for today’s models this instruction is sufficient.

### 10.3.2 Selecting Indicators/Items for a Reflective Measurement

*Reflective measurement* is another term to describe the circumstance where latent variables are assumed to cause observed variables. Observed variables in reflective measurement are called *effect (reflective) indicators*.

- At least three for a unidimensional model; at least two per factor for a multidimensional model (but more is safer).
- The items/indicators should have reasonable internal consistency, they should correlate with each other, and correlate more with themselves than with items on other factors (if multidimensional).
- Negative correlations reduce the reliability of factor measurement, so they should be reverse coded prior to analysis.
- Do not be tempted to specify a factor with indicators that do not measure something. A common mistake is to create a “background” factor and include indicators such as gender, ethnicity, and level of education. *Just what is the predicted relationship between gender and ethnicity?*

## 10.4 CFA Workflow

Below is a screenshot of a CFA workflow. The original document is located in the [Github site](#) that hosts the ReCentering Psych Stats: Psychometrics OER.

Because the intended audience for the ReCentering Psych Stats OER is the scientist-practitioner-advocate, this lesson focuses on the workflow and decisions. As you might guess, the details of CFA can be quite complex and require more investigation and decision-making in models that pose more complexity or empirical challenges.

- Creating an items only dataframe where any items are scaled in the same direction (e.g., negatively worded items are reverse-scored).
- Determining a factor structure that is *identified*, that is
  - A single factor (unidimensional) model has at least three items/indicators
  - Multidimensional models have at least two items per factor
- Specify a series of models, these typically include
  - A unidimensional model (all items on a single factor)
  - A single order structure with correlated factors
  - A second order structure
  - A bifactor structure
- Evaluate model fit with a variety of indicators
  - factor loadings
  - fit indices
- Compare models
- In the event of poor model fit, investigate modification indices and consider respecification

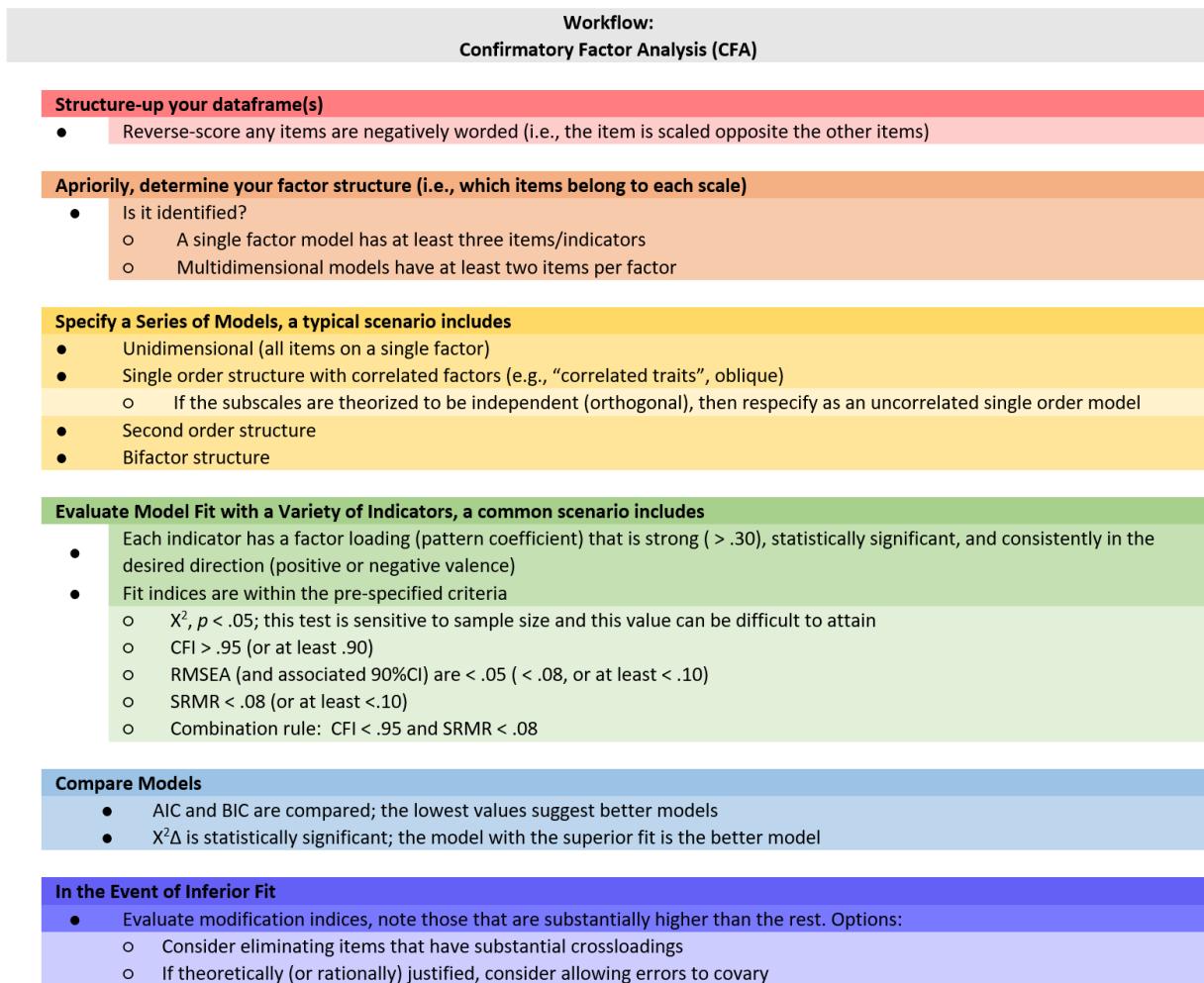


Figure 10.3: Image of a workflow for specifying and evaluating a confirmatory factor analytic model

- eliminating items
- changing factor membership
- allowing errors to covary

#### 10.4.1 CFA in *lavaan* Requires Fluency with the Syntax

- It's really just regression
  - tilda ( $\sim$ , *is regressed on*) is regression operator
  - place DV (y) on left of operator
  - place IVs, separate by + on the right
- f is a latent variable (LV)
- Example:  $y \sim f1 + f2 + x1 + x2$
- LVs must be *defined* by their manifest or latent indicators.
  - the special operator ( $=\sim$ , *is measured/defined by*) is used for this
  - Example:  $f1 =\sim y1 + y2 + y3$
- Variances and covariances are specified with a double tilde operator ( $\sim\sim$ , *is correlated with*)
  - Example of variance:  $y1 \sim\sim y1$  (the relationship with itself)
  - Example of covariance:  $y1 \sim\sim y2$  (relationship with another variable)
  - Example of covariance of a factor:  $f1 \sim\sim f2$

\*Intercepts ( $\sim 1$ ) for observed and LVs are simple, intercept-only regression formulas + Example of variable intercept:  $y1 \sim 1$  + Example of factor intercept:  $f1 \sim 1$

A complete lavaan model is a combination of these formula types, enclosed between single quotation marks. Readibility of model syntax is improved by:

- splitting formulas over multiple lines
- using blank lines within single quote
- labeling with the hashtag

```
myModel <- '# regressions y1 + y2 ~ f1 + f2 + x1 + x2 f1 ~ f2 + f3 f2 ~ f3 + x1 + x2
```

```
# latent variable definitions
f1 =\sim y1 + y2 + y3
f2 =\sim y4 + y5 + y6
f3 =\sim y7 + y8 + y9 + y10

# variances and covariances
y1 ~~ y1
y2 ~~ y2
f1 ~~ f2
```

```
# intercepts
y1 ~ 1
fa ~ 1
```

### 10.4.2 Differing Factor Structures

All models worked in this lesson are *first-order* (or single-order) models; in the next lesson we extend to hierarchical and bifactor models. To provide an advanced cognitive organizer, let's take a look across the models.

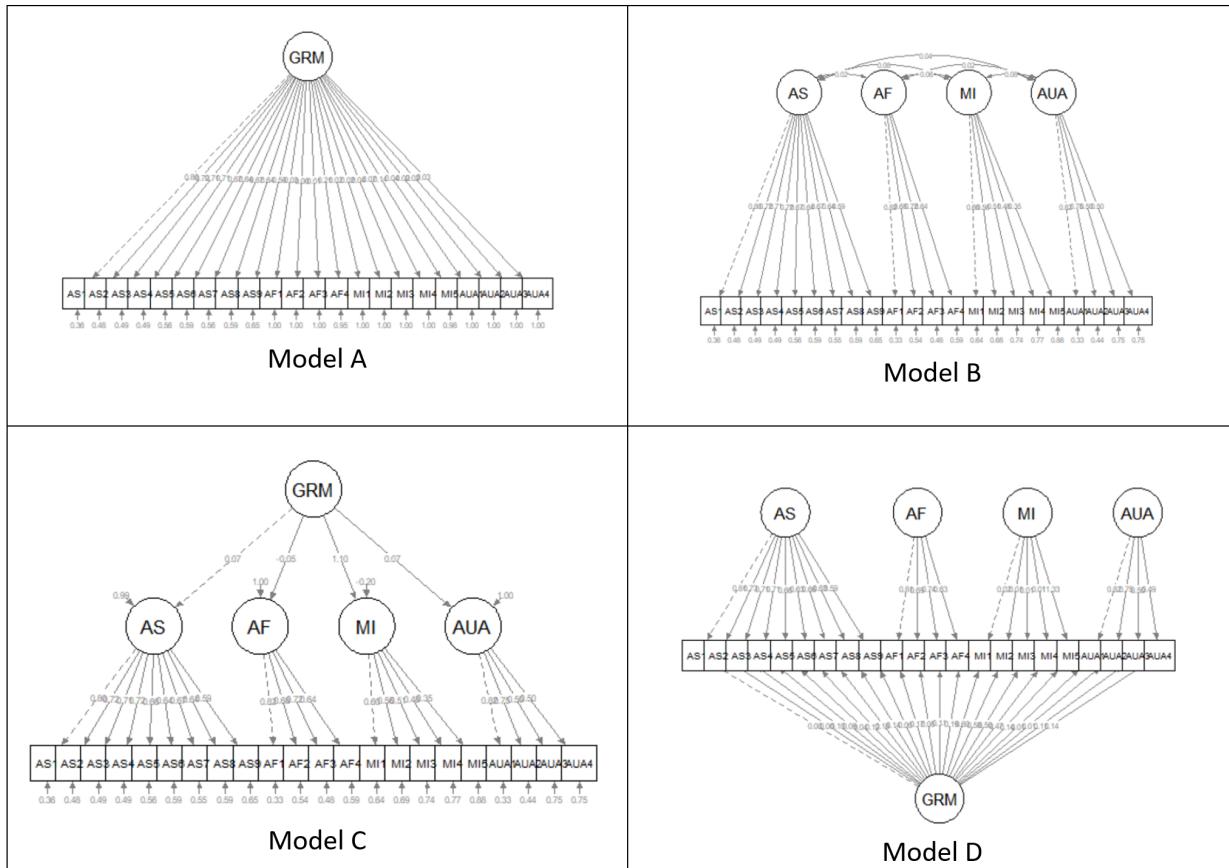


Figure 10.4: Image of first order (uncorrelated and correlated, second order, and bifactor structures)

Models A and B are first-order models. Note that all factors are on a single plane.

- Model A is undimensional, each item is influenced by a single common factor and a term that includes systematic and random error. Note that there is only one *systematic* source of variance for each item and it is from a single source.
- Model B is often referred to as a “correlated traits” model. Here, the larger construct is separated into distinct-yet-correlated elements. The variance of each item is assumed to be a weighted linear function of two or more common factors.

- Models C is a second-order factor structure. Rather than merely being correlated, factors are related because they share a common cause. In this model, the second order factor *explains* why three or more traits are correlated. Note that here is no direct relationship between the item and the target construct. Rather, the relationship between the second-order factor and each item is mediated through the primary factor (yes, an indirect effect!).
- Model D is a bifactor structure. Here each item loads on a general factor. This general factor (bottom row) reflects what is common among the items and represents the individual differences on the target dimension that a researcher is most interested in. Group factors (top row) are now specified as *orthogonal*. The group factors represent common factors measured by the items that explain item response variation not accounted for by the general factor. In some research scenarios, the group factors are termed “nuisance” dimensions. That is, that which they have in common interferes with measuring the primary target of interest.

## 10.5 Research Vignette

This lesson’s research vignette emerges from Keum et al’s Gendered Racial Microaggressions Scale for Asian American Women (GRMSAAW; [Keum et al., 2018]). The article reports on two separate studies that comprised the development, refinement, and psychometric evaluation of two, parallel, versions (stress appraisal, frequency) of scale. I simulated data from the final construction of the frequency version as the basis of the lecture. If the scale looks somewhat familiar it is because the authors used the Gendered Racial Microaggressions Scale for Black Women [Lewis and Neville, 2015] as a model.

Keum et al. [2018] reported support for a total scale score (22 items) and four subscales. Below, I list the four subscales, their number of items, and a single example item. At the outset, let me provide a content advisory For those who hold this particular identity (or related identities) the content in the items may be upsetting. In other lessons, I often provide a variable name that gives an indication of the primary content of the item. In the case of the GRMSAAW, I will simply provide an abbreviation of the subscale name and its respective item number. This will allow us to easily inspect the alignment of the item with its intended factor, and hopefully minimize discomfort. If you are not a member of this particular identity, I encourage you to learn about these microaggressions by reading the article in its entirety. Please do not ask members of this group to explain why these microaggressions are harmful or ask if they have encountered them.

There are 22 items on the GRMSAAW scale. The frequency scaling ranged included: 0(*never*), 1 (*rarely*), 2(*sometimes*), 3(*often*), 4(*very frequently*), and 5(*always*).

The four factors, number of items, and sample item are as follows:

- Ascribed Submissiveness
  - 9 items
  - “Others have been surprised when I disagree with them.”
  - Abbreviated in the simulated data as “AS#”
- Asian Fetishism
  - 4 items
  - “Others have treated me as if I am always open to sexual advances.”

- Abbreviated in the simulated data as “AF#”
- Media Invalidation
  - 5 items
  - “I see AAW playing the same type of characters (e.g., Kung Fu woman, sidekick, mistress, tiger mom) in the media.”
  - Abbreviated in the simulated data as “MI#”
- Assumptions of Universal Appearance
  - 4 items
  - “Others have pointed out physical traits in AAW that do not look ‘Asian’”
  - Abbreviated in the simulated data as “AUA#”

Four additional scales were reported in the Keum et al. article [Keum et al., 2018]. Fortunately, I was able to find factor loadings from the original psychometric article or subsequent publications. For multidimensional scales, I assign variable names according to the scale to which the item belongs (e.g., Env42). In contrast, when subscales or short unidimensional scales were used, I assigned variable names based on item content (e.g., “blue”). In my own work, I prefer item-level names so that I can quickly see (without having to look up the item names) how the items are behaving. While the focus of this series of chapters is on the Gendered Racial Microaggressions Scale for Asian American Women scale, this simulated data might be useful to you in one or more of the suggestions for practice (e.g., examining the psychometric characteristics of one or the other scales). The scales, their original citation, and information about how I simulated data for each are listed below.

- **Racial Microaggressions Scale** (RMAS; [Torres-Harding et al., 2012]) is a 32-item scale with Likert scaling ranging from 0 (*never*) to 3 (*often/frequent*). Higher scores represent greater frequency of perceived microaggressions. I simulated data at the subscale level. The RMAS has six subscales, but only four (Invisibility, Low-Achieving/Undesirable Culture, Foreigner/Not Belonging, and Environmental Invalidation) were used in the study. Data were simulated using factor loadings (from the four factors) in the source article.
- **Schedule of Sexist Events** (SSE; [Klonoff and Landrine, 1995]) is a 20-item scale that with Likert scaling ranging from 1 (*the event has never happened to me*) to 6 (*the event happened almost all [i.e., more than 70%] of the time*). Higher scores represent greater frequency of everyday sexist events. I simulated data the subscale level. Within two larger scales (recent events, lifetime events), there are three subscales: Sexist Degradation and Its Consequences, Unfair/Sexist Events at Work/School, and Unfair Treatment in Distant and Close Relationships. Data were simulated using factor loadings from the source article.
- **PHQ-9** [Kroenke et al., 2001] is a 9-item scale with Likert scaling ranging from 0 (*not at all*) to 3 (*nearly every day*). Higher scores indicate higher levels of depression. I simulated data by estimating factor loadings from Brattmyr et al. [2022].
- **Internalized Racism in Asian American Scale** (IRAAS [Choi et al., 2017]) is a 14-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Higher scores indicate greater internalized racism. Data were simulated using the factor loadings from the bifactor model in the source article.

As you consider homework options, there is sufficient simulated data to use the RMAS, SSE, or IRAAS.

Below, I walk through the data simulation. This is not an essential portion of the lesson, but I will lecture it in case you are interested. None of the items are negatively worded (relative to the other items), so there is no need to reverse-score any items.

Simulating the data involved using factor loadings, means, standard deviations, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

```
#Entering the intercorrelations, means, and standard deviations from the journal article

Keum_GRMS_generating_model <- '
  #measurement model
  General =~ .50*AS1 + .44*AS2 + .50*AS3 + .33*AS4 + .58*AS5 + .49*AS6 + .51*AS7 + .53*AS8
  AS =~ .68*AS1 + .65*AS2 + .53*AS3 + .55*AS4 + .54*AS5 + .55*AS6 + .42*AS7 + .47*AS8 + .49*AS9
  AF =~ .63*AF1 + .45*AF2 + .56*AF3 + .54*AF4
  AUA =~ .55*AUA1 + .55*AUA2 + .31*AUA3 + .31*AUA4
  MI =~ .27*MI1 + .53*MI2 + .57*MI3 + .29*MI4 + .09*MI5
  RMAS_FOR =~ .66*FOR1 + .90*FOR2 + .63*FOR4
  RMAS_LOW =~ .64*LOW22 + .54*LOW23 + .49*LOW28 + .63*LOW29 + .58*LOW30 + .67*LOW32 + .67*LOW33
  RMAS_INV =~ .66*INV33 + .70*INV39 + .79*INV40 + .71*INV41 + .71*INV47 + .61*INV49 + .61*INV50
  RMAS_ENV =~ .71*ENV42 + .70*ENV43 + .74*ENV44 + .57*ENV45 + .54*ENV46

  SSEL_Deg =~ .77*LDeg18 + .73*LDeg19 + .71*LDeg21 + .71*LDeg15 + .67*LDeg16 + .67*LDeg17
  SSEL_dRel =~ .69*LdRel4 + .68*LdRel6 + .64*LdRel7 + .64*LdRel5 + .63*LdRel11 + .49*LdRel12
  SSEL_cRel =~ .73*LcRel11 + .68*LcRel9 + .66*LcRel23
  SSEL_Work =~ .73*LWork17 + .10*LWork10 + .64*LWork2

  SSER_Deg =~ .72*RDeg15 + .71*RDeg21 + .69*RDeg18 + .68*RDeg16 + .68*RDeg13 + .65*RDeg12
  SSER_dRel =~ .74*RDeg4 + .67*RDeg6 + .64*RDeg5 + .54*RDeg7 + .51*RDeg1
  SSER_cRel =~ .69*RcRel9 + .59*RcRel11 + .53*RcRel23
  SSER_Work =~ .72*RWork10 + .67*RWork2 + .62*RWork17 + .51*RWork3

  SSE_Lifetime =~ SSEL_Deg + SSEL_dRel + SSEL_cRel + SSEL_Work
  SSE_Recent =~ SSER_Deg + SSER_dRel + SSEL_cRel + SSER_Work

  PHQ9 =~ .798*anhedonia + .425*down + .591*sleep + .913*lo_energy + .441*appetite +
    .441*anorexia + .441*restlessness + .441*fatigue + .441*headache + .441*nausea + .441*constipation +
    .441*urinary_symptoms + .441*sexual_symptoms + .441*depression + .441*suicidality + .441*other_symptoms

  gIRAAS =~ .51*SN1 + .69*SN2 + .63*SN3 + .65*SN4 + .67*WS5 + .60*WS6 + .74*WS7 + .44*WS8
  SelfNegativity =~ .60*SN1 + .50*SN2 + .63*SN3 + .43*SN4
  WeakStereotypes =~ .38*WS5 + .22*WS6 + .10*WS7 + .77*WS8 + .34*WS9 + .14*WS10
  AppearanceBias =~ .38*AB11 + .28*AB12 + .50*AB13 + .18*AB14

  #Means
  #Keum et al reported total scale scores, I divided those totals by the number of items
  AS ~ 3.25*1
  AF ~ 3.34*1
  AUA ~ 4.52
```

```

MI ~ 5.77*1
General ~ 3.81*1
RMAS_FOR ~ 3.05*1
RMAS_LOW ~ 2.6*1
RMAS_INV ~ 2.105*1
RMAS_ENV ~ 3.126*1
SSEL_Deg ~ 2.55*1
SSEL_dRel ~ 1.96*1
SSEL_cRel ~ 3.10*1
SSEL_Work ~ 1.66*1
SSER_Deg ~ 2.02*1
SSER_dRel ~ 1.592*1
SSER_cRel ~ 1.777*1
SSER_Work ~ 1.3925*1
SSER_Lifetime ~ 2.8245*1
SSER_Recent ~ 2.4875*1
PHQ9 ~ 1.836*1
gIRAAS ~ 2.246*1

```

```

#Correlations
AS ~~ .00*AF
AS ~~ .00*AUA
AS ~~ .00*MI
AS ~~ .00*General
AS ~~ .28*RMAS_FOR
AS ~~ .24*RMAS_LOW
AS ~~ .46*RMAS_INV
AS ~~ .16*RMAS_ENV
AS ~~ .40*SSE_Lifetime
AS ~~ .28*SSE_Recent
AS ~~ .15*PHQ9
AS ~~ .13*gIRAAS

```

```

AF ~~ .00*AUA
AF ~~ .00*MI
AF ~~ .00*General
AF ~~ .02*RMAS_FOR
AF ~~ .05*RMAS_LOW
AF ~~ .11*RMAS_INV
AF ~~ .07*RMAS_ENV
AF ~~ .34*SSE_Lifetime
AF ~~ .27*SSE_Recent
AF ~~ -.04*PHQ9
AF ~~ .21*gIRAAS

```

```

AUA ~~ .00*MI
AUA ~~ .00*General

```

```

AUA ~~ .18*RMAS_FOR
AUA ~~ .20*RMAS_LOW
AUA ~~ .01*RMAS_INV
AUA ~~ -.04*RMAS_ENV
AUA ~~ .02*SSE_Lifetime
AUA ~~ .92*SSE_Recent
AUA ~~ .02*PHQ9
AUA ~~ .17*gIRAAS

MI ~~ .00*General
MI ~~ -.02*RMAS_FOR
MI ~~ .08*RMAS_LOW
MI ~~ .31*RMAS_INV
MI ~~ .36*RMAS_ENV
MI ~~ .15*SSE_Lifetime
MI ~~ .08*SSE_Recent
MI ~~ -.05*PHQ9
MI ~~ -.03*gIRAAS

General ~~ .34*RMAS_FOR
General ~~ .63*RMAS_LOW
General ~~ .44*RMAS_INV
General ~~ .45*RMAS_ENV
General ~~ .54*SSE_Lifetime
General ~~ .46*SSE_Recent
General ~~ .31*PHQ9
General ~~ -.06*gIRAAS

RMAS_FOR ~~ .57*RMAS_LOW
RMAS_FOR ~~ .56*RMAS_INV
RMAS_FOR ~~ .37*RMAS_ENV
RMAS_FOR ~~ .33*SSE_Lifetime
RMAS_FOR ~~ .25*SSE_Recent
RMAS_FOR ~~ .10*PHQ9
RMAS_FOR ~~ .02*gIRAAS

RMAS_LOW ~~ .69*RMAS_INV
RMAS_LOW ~~ .48*RMAS_ENV
RMAS_LOW ~~ .67*SSE_Lifetime
RMAS_LOW ~~ .57*SSE_Recent
RMAS_LOW ~~ .30*PHQ9
RMAS_LOW ~~ .16*gIRAAS

RMAS_INV ~~ .59*RMAS_ENV
RMAS_INV ~~ .63*SSE_Lifetime
RMAS_INV ~~ .52*SSE_Recent

```

```

RMAS_INV ~~ .32*PHQ9
RMAS_INV ~~ .23*gIRAAS

RMAS_ENV ~~ .46*SSE_Lifetime
RMAS_ENV ~~ .31*SSE_Recent
RMAS_ENV ~~ .11*PHQ9
RMAS_ENV ~~ .07*gIRAAS

SSE_Lifetime ~~ .83*SSE_Recent
SSE_Lifetime ~~ .30*PHQ9
SSE_Lifetime ~~ .14*gIRAAS

SSE_Recent ~~ .30*PHQ9
SSE_Recent ~~ .20*gIRAAS

PHQ9 ~~ .18*gIRAAS

#Correlations between SES scales from the Klonoff and Landrine article
#Note that in the article the factor orders were reversed
SSEL_Deg ~~ .64*SSEL_dRel
SSEL_Deg ~~ .61*SSEL_cRel
SSEL_Deg ~~ .50*SSEL_Work
SSEL_dRel ~~ .57*SSEL_cRel
SSEL_dRel ~~ .57*SSEL_Work
SSEL_cRel ~~ .47*SSEL_Work

SSER_Deg ~ .54*SSER_dRel
SSER_Deg ~ .54*SSER_Work
SSER_Deg ~ .59*SSER_cRel
SSER_dRel ~ .56*SSER_Work
SSER_dRel ~ .46*SSER_cRel
SSER_Work ~ .43*SSER_cRel

SSE_Lifetime ~ .75*SSE_Recent

'

set.seed(240311)
dfGRMSAAW <- lavaan::simulateData(model = Keum_GRMS_generating_model,
                                     model.type = "sem",
                                     meanstructure = T,
                                     sample.nobs=304,
                                     standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(dfGRMSAAW))

```

```

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 22 are the GRMS items
#Rows 23 thru 47 are the RMAS
#Rows 48 thru 87 are the SSE
#Rows 88 thru 96 are the PHQ9
#Rows 97 thru 110 are the IRAAS
#Rows 111 thru 112 are scale scores for SSE

for(i in 1:ncol(dfGRMSAAW)){
  if(i >= 1 & i <= 22){
    dfGRMSAAW[,i] <- scales::rescale(dfGRMSAAW[,i], c(1, 5))
  }
  if(i >= 23 & i <= 47){
    dfGRMSAAW[,i] <- scales::rescale(dfGRMSAAW[,i], c(0, 3))
  }
  if(i >= 48 & i <= 87){
    dfGRMSAAW[,i] <- scales::rescale(dfGRMSAAW[,i], c(1, 6))
  }
  if(i >= 88 & i <= 96){
    dfGRMSAAW[,i] <- scales::rescale(dfGRMSAAW[,i], c(0, 3))
  }
  if(i >= 97 & i <= 110){
    dfGRMSAAW[,i] <- scales::rescale(dfGRMSAAW[,i], c(1, 6))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
dfGRMSAAW <- dfGRMSAAW %>% round(0)

#quick check of my work
#psych::describe(dfGRMSAAW)

```

The optional script below will let you save the simulated data to your computing environment as either a .csv file (think “Excel lite”) or .rds object (preserves any formatting you might do). If you save the .csv file and bring it back in, you will lose any formatting (e.g., ordered factors will be interpreted as character variables).

```

#write the simulated data as a .csv
#write.table(dfGRMSAAW, file="dfGRMSAAW.csv", sep=",", col.names=TRUE, row.names=FALSE)
#bring back the simulated dat from a .csv file
#dfGRMSAAW <- read.csv ("dfGRMSAAW.csv", header = TRUE)

```

An .rds file preserves all formatting to variables prior to the export and re-import. For the purpose of this chapter, you don’t need to do either. That is, you can re-simulate the data each time you work the problem.

```
#to save the df as an .rds (think "R object") file on your computer; it should save in the same
#saveRDS(dfGRMSAAW, "dfGRMSAAW.rds")
#bring back the simulated dat from an .rds file
#dfGRMSAAW <- readRDS("dfGRMSAAW.rds")
```

### 10.5.1 Modeling the GRMSAAW as Unidimensional

Let's start simply, taking the GRMSAAW data and seeing about its fit as a unidimensional instrument. First evaluating multi-dimensional measures as unidimensional is a common practice. And there are two reasons:

- Operationally, it's a check to see that data, script, and so forth. are all working.
- If you can't reject a single-factor model (e.g., if there is a strong support for such), then it makes little sense to evaluate models with more factors [Kline, 2016].

With a single factor model:

- GRMSAAW is a latent variable and can be named anything. We know this because it is followed by: =~
- All the items follow and are “added” with the plus sign
  - Don't let this fool you...the assumption behind SEM/CFA is that the LV *causes* the score on the item/indicator. Recall, item/indicator scores are influenced by the LV and error.
- The entire model is enclosed in tic marks (‘ and ’)

```
grmsAAWmod1 <- 'GRMSAAW =~ AS1 + AS2 + AS3 + AS4 + AS5 + AS6 + AS7 + AS8 + AS9 + AF1 + AF2 + AF3 + AF4 + AF5'
```

```
[1] "GRMSAAW =~ AS1 + AS2 + AS3 + AS4 + AS5 + AS6 + AS7 + AS8 + AS9 + AF1 + AF2 + AF3 + AF4 + AF5"
```

The object representing the model is then included in the *lavaan::cfa()* along with the dataset.

We can ask for a summary of the object representing the results.

```
grmsAAW1fit <- lavaan::cfa (grmsAAWmod1, data = dfGRMSAAW)
lavaan::summary(grmsAAW1fit, fit.measures=TRUE, standardized=TRUE, rsquare = TRUE)
```

*lavaan* 0.6.17 ended normally after 33 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	44
Number of observations	304

Model Test User Model:

Test statistic	466.669
Degrees of freedom	209
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	1479.910
Degrees of freedom	231
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.794
Tucker-Lewis Index (TLI)	0.772

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-7106.670
Loglikelihood unrestricted model (H1)	-6873.336
Akaike (AIC)	14301.341
Bayesian (BIC)	14464.890
Sample-size adjusted Bayesian (SABIC)	14325.344

Root Mean Square Error of Approximation:

RMSEA	0.064
90 Percent confidence interval - lower	0.056
90 Percent confidence interval - upper	0.071
P-value H_0: RMSEA <= 0.050	0.002
P-value H_0: RMSEA >= 0.080	0.000

Standardized Root Mean Square Residual:

SRMR	0.068
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
GRMSAAW =~						
AS1	1.000				0.416	0.545

AS2	0.994	0.143	6.943	0.000	0.413	0.504
AS3	0.974	0.137	7.101	0.000	0.405	0.520
AS4	0.893	0.129	6.914	0.000	0.371	0.501
AS5	1.198	0.152	7.897	0.000	0.498	0.608
AS6	0.701	0.110	6.401	0.000	0.291	0.452
AS7	0.892	0.130	6.891	0.000	0.371	0.499
AS8	0.974	0.137	7.093	0.000	0.405	0.519
AS9	0.837	0.117	7.179	0.000	0.348	0.528
AF1	0.712	0.120	5.922	0.000	0.296	0.409
AF2	1.016	0.138	7.360	0.000	0.422	0.547
AF3	0.622	0.117	5.324	0.000	0.258	0.359
AF4	0.769	0.134	5.716	0.000	0.319	0.392
MI1	0.590	0.109	5.401	0.000	0.245	0.366
MI2	0.661	0.122	5.417	0.000	0.275	0.367
MI3	0.831	0.139	5.992	0.000	0.345	0.415
MI4	0.647	0.133	4.874	0.000	0.269	0.324
MI5	0.604	0.118	5.099	0.000	0.251	0.342
AUA1	0.726	0.138	5.258	0.000	0.302	0.354
AUA2	0.889	0.132	6.732	0.000	0.370	0.483
AUA3	0.702	0.116	6.075	0.000	0.292	0.422
AUA4	0.904	0.132	6.835	0.000	0.376	0.493

## Variances:

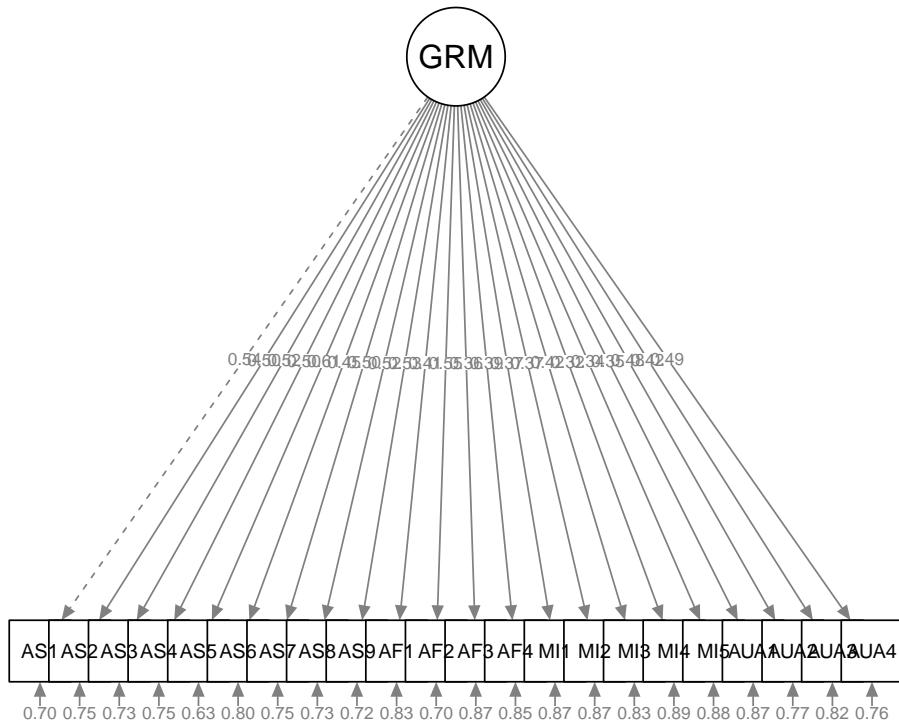
	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.AS1	0.410	0.036	11.476	0.000	0.410	0.703
.AS2	0.502	0.043	11.641	0.000	0.502	0.746
.AS3	0.443	0.038	11.580	0.000	0.443	0.730
.AS4	0.411	0.035	11.652	0.000	0.411	0.749
.AS5	0.423	0.038	11.143	0.000	0.423	0.631
.AS6	0.331	0.028	11.810	0.000	0.331	0.796
.AS7	0.416	0.036	11.660	0.000	0.416	0.751
.AS8	0.445	0.038	11.584	0.000	0.445	0.731
.AS9	0.313	0.027	11.548	0.000	0.313	0.721
.AF1	0.436	0.037	11.922	0.000	0.436	0.833
.AF2	0.418	0.036	11.465	0.000	0.418	0.701
.AF3	0.450	0.037	12.029	0.000	0.450	0.871
.AF4	0.563	0.047	11.963	0.000	0.563	0.847
.MI1	0.390	0.032	12.017	0.000	0.390	0.866
.MI2	0.485	0.040	12.014	0.000	0.485	0.865
.MI3	0.572	0.048	11.908	0.000	0.572	0.828
.MI4	0.615	0.051	12.092	0.000	0.615	0.895
.MI5	0.477	0.040	12.062	0.000	0.477	0.883
.AUA1	0.635	0.053	12.039	0.000	0.635	0.875
.AUA2	0.449	0.038	11.714	0.000	0.449	0.767
.AUA3	0.392	0.033	11.890	0.000	0.392	0.821
.AUA4	0.439	0.038	11.679	0.000	0.439	0.757
GRMSAAW	0.173	0.036	4.795	0.000	1.000	1.000

R-Square:

	Estimate
AS1	0.297
AS2	0.254
AS3	0.270
AS4	0.251
AS5	0.369
AS6	0.204
AS7	0.249
AS8	0.269
AS9	0.279
AF1	0.167
AF2	0.299
AF3	0.129
AF4	0.153
MI1	0.134
MI2	0.135
MI3	0.172
MI4	0.105
MI5	0.117
AUA1	0.125
AUA2	0.233
AUA3	0.179
AUA4	0.243

I find it helpful to immediately plot what we did. A quick look alerts me to errors.

```
semPlot::semPaths(grmsAAW1fit, layout = "tree", style = "lisrel", what = "col", whatLabels = "
```



### 10.5.1.1 Interpreting the Output

With a quick look at the plot, let's work through the results. Rosseel's (2019) *lavaan* tutorial is a useful resource in walking through the output.

The *header* is the first few lines of the information. It contains:

- the *lavaan* version number (0.6-9 that I'm using on 10/4/2021)
- maximum likelihood (ML) was used as the estimator
- confirmation that the specification converged normally after 28 iterations
- 304 cases were used in this analysis (would be less if some were skipped because of missing data)
- the model user test statistic, df, and corresponding p value:  $\chi^2(209) = 1004.136, p < .001$

**Fit statistics** are included in the second section. They are only shown when the argument “*fit.measures* = TRUE” is in the script. Standardized values are not the default, they require the argument, “*standardized* = TRUE”. We'll come back to these shortly...

*Parameter estimates* is the last section.

For now we are interested in the Latent Variables section.

- *Estimate* contains the estimated or fixed parameter value for each model parameter;
- *Std. err* is the standard error for each estimated parameter;
- *Z-value* is the Wald statistic (the parameter divided by its SE)
- $P(>/z|)$  is the p value for testing the null hypothesis that the parameter equals zero in the population
- *Std.lv* standardizes only the LVs
- *Std.all* both latent and observed variables are standardized; this is considered the “completely standardized solution”

Note that item AS1 might seem incomplete – there is only a 1.000 and a value for the Std.lv. Recall we used this to scale the single factor by fixing its value to 1.000. Coefficients that are fixed to 1.0 to scale a factor have no standard errors and therefore no significance test.

The SE and associated  $p$  values are associated with the unstandardized estimates. Intuitively, it is easiest for me to understand the relative magnitude of the pattern coefficients by looking at the *Std.all* column. We can see that the items associated with what we will soon define as the AS factor are all strong and positive. The remaining items have variable loadings with many of the being quite low, non-significant, and even negatively valenced.

Let's examine to the middle set metrics which assess *global fit*.

CFA falls into a *modeling* approach to evaluating results. While it provides some flexibility (we get away from the strict, NHST approach of  $p < .05$ ) there is greater interpretive ambiguity.

Fit statistics tend to be clustered together based on their approach to summarizing the *goodness* or *badness* of fit.

#### 10.5.1.2 Model Test *User Model*:

The chi-square statistic that evaluates the *exact-fit hypothesis* that there is no difference between the covariances predicted by the model, given the parameter estimates, and the population covariance matrix. Rejecting the hypothesis says that,

- the data contain covariance information that speak against the model, and
- the researcher should explain model-data discrepancies that exceed those expected by sampling error.

Traditional interpretation of the chi-square is an *accept-support test* where the null hypothesis represents the researchers' belief that the model is correct. This means that the absence of statistical significance ( $p > .05$ ) that supports the model. This is backwards from our usual *reject-support test* approach.

The  $\chi^2$  is frequently criticized:

- *accept-support test* approaches are logically weaker because the failure to disprove an assertion (the exact-fit hypothesis) does not prove that the assertion is true;

- too small a sample size (low power) makes it more likely that the model will be retained;
- CFA/SEM, though, requires large samples and so the  $\chi^2$  is frequently statistically significant – which rejects the researchers' model;

Kline [2016] recommends that we treat the  $\chi^2$  like a smoke alarm – if the alarm sounds, there may or may not be a fire (a serious model-data discrepancy), but we should treat the alarm seriously and further inspect issues of fit.

For our unidimensional GRMSAAW CFA  $\chi^2(209) = 1004.136, p < .001$ , this significant value is not what we want because it says that our specified model is different than the covariances in the model.

### 10.5.1.3 Model Test *Baseline* Model

This model is the *independence* model. That is, there is complete independence of all variables in the model (i.e., in which all correlations among variables are zero). This is the most restricted model. It is typical for chi-square values to be quite high (as it is in our example: 2114.899). On its own, this model is not useful to us. It is used, though, in comparisons of *incremental fit*.

### 10.5.1.4 Incremental Fit Indices (User versus Baseline Models)

Incremental fit indices ask the question, how much better is the fit of our specified model to the data than the baseline model (where it is assumed no relations between the variables).

The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are *goodness of fit* statistics, ranging from 0 to 1.0 where 1.0 is best.

**CFI:** compares the amount of departure from close fit for the researcher's model against that of the independence/baseline (null) model.

$$CFI = 1 - \frac{\hat{\Delta}_M}{\hat{\Delta}_B}$$

We can actually calculate this using the baseline and chi-square values from our own data:

```
1 - (1004.136/2114.899)
```

```
[1] 0.5252085
```

Where there is no departure from close fit, then CFI will equal 1.0. We interpret the value of the CFI as a percent of how much better the researcher's model is than the baseline model. While 58% sounds like an improvement – Hu and Bentler (1999) stated that “acceptable fit” is achieved when the  $CFI \geq .95$  and  $SRMR \leq .08$ ; the **combination rule**. It is important to note that later simulation studies have not supported those thresholds.

**TLI:** aka the **non-normed fit index (NNFI)** controls for  $df_M$  from the researcher's model and  $df_B$  from the baseline model. As such, it imposes a greater relative penalty for model complexity than the CFI. The TLI is a bit unstable in that the values can exceed 1.0.

Because the two measures are so related, only one should be reported (I typically see the CFI).

For our unidimensional GRMSAAW CFA,  $CFI = .578$  and  $TLI = .534$ . While these predict around 58% better than the baseline/independence model, it does not come close to the standard of  $\geq .95$ .

*I note that our hand calcuation of user and baseline models did not result in the exact CFI. I do not know why.*

#### 10.5.1.5 Loglikelihood and Information Criteria

The **Aikaike Information Criterion (AIC)** and the **Bayesian Information Criterion (BIC)** utilize an information theory approach to data analysis by combining statistical estimation and model selection into a single framework. The BIC augments the AIC by taking sample size into consideration.

The AIC and BIC are usually used to select among competing nonhierarchical models and are only used in comparison with each other. Thus our current values of 17755.028 (AIC) and 17918.577 (BIC) are meaningless on their own. The model with the smallest value of the predictive fit index is chosen as the one that is most likely to replicate. It means that this model has relatively better fit and fewer free parameters than competing models.

For our unidimensional GRMSAAW CFA we'll return to these values to compare a correlated, four-factor solution.

#### 10.5.1.6 Root Mean Square Error of Approximation

The RMSEA is an absolute fit index scaled as a *badness-of-fit* statistic where a value of 0.00 is the best fit. The RMSEA favors models with more degrees of freedom and larger sample sizes. A unique aspect of the RMSEA is its 90% confidence interval.

While there is chatter/controversy about what constitutes an acceptable value, there is general consensus that  $RMSEA \geq .10$  points to serious problems. An  $RMSEA \leq .05$  is desired. Watching the upper bound of the confidence interval is important to see that it isn't sneaking into the danger zone.

For our unidimensional GRMSAAW CFA,  $RMSEA = .112$ , 90% CI(.105, .119). Unfortunately this value points to serious problems.

#### 10.5.1.7 Standardized Root Mean Square Residual

The SRMR is an absolute fit index that is a *badness-of-fit* statistic (i.e., perfect model fit is when the value = 0.00 and increasingly higher values indicate the “badness”).

The SRMR is a standardized version of the **root mean square residual (RMR)**, which is a measure of the mean absolute covariance residual. Standardizing the value facilitates interpretation.

Poor fit is indicated when  $SRMR \geq .10$ .

Recall, Hu and Bentler's **combination rule** (which is somewhat contested) suggested that the SRMR be interpreted along with the CFI such that:  $CFI \geq .95$  and  $SRMR \leq .05$ .

For our unidimensional GRMSAAW CFA,  $SRMR = .124$ . Not good.

Inspecting the residuals (we look for relatively large values) may help understand the source of poor fit, so let's do that.

```
lavaan::fitted(grmsAAW1fit)
```

\$cov	AS1	AS2	AS3	AS4	AS5	AS6	AS7	AS8	AS9	AF1	AF2	AF3
AS1	0.582											
AS2	0.172	0.672										
AS3	0.168	0.167	0.607									
AS4	0.154	0.153	0.150	0.548								
AS5	0.207	0.206	0.202	0.185	0.671							
AS6	0.121	0.120	0.118	0.108	0.145	0.416						
AS7	0.154	0.153	0.150	0.138	0.185	0.108	0.553					
AS8	0.168	0.167	0.164	0.150	0.201	0.118	0.150	0.608				
AS9	0.144	0.144	0.141	0.129	0.173	0.101	0.129	0.141	0.434			
AF1	0.123	0.122	0.120	0.110	0.147	0.086	0.110	0.120	0.103	0.524		
AF2	0.175	0.174	0.171	0.157	0.210	0.123	0.157	0.171	0.147	0.125	0.596	
AF3	0.107	0.107	0.105	0.096	0.129	0.075	0.096	0.105	0.090	0.077	0.109	0.517
AF4	0.133	0.132	0.129	0.118	0.159	0.093	0.118	0.129	0.111	0.095	0.135	0.083
MI1	0.102	0.101	0.099	0.091	0.122	0.071	0.091	0.099	0.085	0.073	0.104	0.063
MI2	0.114	0.113	0.111	0.102	0.137	0.080	0.102	0.111	0.096	0.081	0.116	0.071
MI3	0.143	0.143	0.140	0.128	0.172	0.101	0.128	0.140	0.120	0.102	0.146	0.089
MI4	0.112	0.111	0.109	0.100	0.134	0.078	0.100	0.109	0.093	0.080	0.113	0.069
MI5	0.104	0.104	0.102	0.093	0.125	0.073	0.093	0.102	0.087	0.074	0.106	0.065
AUA1	0.125	0.125	0.122	0.112	0.150	0.088	0.112	0.122	0.105	0.089	0.127	0.078
AUA2	0.154	0.153	0.150	0.137	0.184	0.108	0.137	0.150	0.128	0.109	0.156	0.096
AUA3	0.121	0.121	0.118	0.108	0.145	0.085	0.108	0.118	0.101	0.086	0.123	0.075
AUA4	0.156	0.155	0.152	0.139	0.187	0.109	0.139	0.152	0.131	0.111	0.159	0.097
	AF4	MI1	MI2	MI3	MI4	MI5	AUA1	AUA2	AUA3	AUA4		
AS1												
AS2												
AS3												
AS4												
AS5												
AS6												
AS7												
AS8												
AS9												
AF1												
AF2												
AF3												
AF4	0.665											
MI1	0.078	0.450										
MI2	0.088	0.067	0.561									
MI3	0.110	0.085	0.095	0.691								
MI4	0.086	0.066	0.074	0.093	0.688							

```

MI5  0.080 0.062 0.069 0.087 0.067 0.540
AUA1 0.096 0.074 0.083 0.104 0.081 0.076 0.726
AUA2 0.118 0.091 0.102 0.128 0.099 0.093 0.112 0.585
AUA3 0.093 0.072 0.080 0.101 0.078 0.073 0.088 0.108 0.477
AUA4 0.120 0.092 0.103 0.130 0.101 0.094 0.113 0.139 0.110 0.580

```

```

#lavaan::residuals(grmsAAW1fit, type = "raw")
#lavaan::residuals(grmsAAW1fit, type = "standardized")

#will hashtag out for knitted file
lavaan::residuals(grmsAAW1fit, type = "cor")

```

```

$type
[1] "cor.bollen"

$cov
      AS1    AS2    AS3    AS4    AS5    AS6    AS7    AS8    AS9    AF1
AS1  0.000
AS2  0.135  0.000
AS3  0.070  0.060  0.000
AS4  0.057  0.124  0.061  0.000
AS5  0.095  0.023 -0.035  0.024  0.000
AS6  0.029  0.112  0.018  0.060  0.121  0.000
AS7  0.012  0.071  0.043  0.013 -0.032 -0.068  0.000
AS8  0.016  0.074  0.002  0.021  0.016  0.007  0.090  0.000
AS9  0.014  0.126  0.023  0.104  0.097  0.092 -0.011  0.014  0.000
AF1 -0.047 -0.106  0.026  0.004 -0.046 -0.099 -0.021  0.028 -0.079  0.000
AF2 -0.057 -0.116  0.002  0.014 -0.057 -0.051 -0.057 -0.045 -0.053  0.171
AF3 -0.087 -0.111  0.021 -0.062 -0.057 -0.037  0.046  0.040 -0.082  0.264
AF4 -0.134 -0.046  0.027 -0.056 -0.021 -0.067 -0.046 -0.086  0.006  0.212
MI1 -0.044 -0.058 -0.075 -0.057 -0.068 -0.045 -0.032 -0.070 -0.135  0.072
MI2 -0.052 -0.062 -0.087 -0.071 -0.012  0.015  0.023  0.028 -0.100 -0.056
MI3 -0.061 -0.077 -0.058 -0.039 -0.011 -0.003 -0.033  0.046 -0.017 -0.051
MI4 -0.098 -0.111  0.034 -0.115 -0.081 -0.034  0.043 -0.088 -0.029 -0.010
MI5  0.049 -0.027 -0.045  0.036  0.033 -0.009 -0.030 -0.031 -0.069 -0.033
AUA1 -0.041 -0.108 -0.046 -0.101 -0.078  0.015  0.017 -0.046 -0.117 -0.112
AUA2 -0.109 -0.050 -0.023 -0.051 -0.019 -0.067 -0.020 -0.030 -0.016 -0.012
AUA3  0.052 -0.075 -0.026 -0.064 -0.060 -0.093 -0.016 -0.080  0.027 -0.058
AUA4 -0.007 -0.069 -0.067 -0.127  0.033 -0.069  0.007 -0.008 -0.046  0.028
      AF2    AF3    AF4    MI1    MI2    MI3    MI4    MI5    AUA1    AUA2
AS1
AS2
AS3
AS4
AS5
AS6
AS7

```

```

AS8
AS9
AF1
AF2 0.000
AF3 0.125 0.000
AF4 0.161 0.116 0.000
MI1 0.101 0.048 0.043 0.000
MI2 -0.034 0.023 -0.007 0.102 0.000
MI3 -0.002 -0.056 -0.037 0.107 0.202 0.000
MI4 0.050 0.034 0.088 0.129 0.074 0.112 0.000
MI5 0.010 0.008 0.026 -0.070 0.101 -0.013 0.048 0.000
AUA1 0.086 -0.026 -0.060 0.092 0.078 0.110 0.039 0.034 0.000
AUA2 0.004 -0.017 0.013 0.038 0.077 0.048 0.019 0.034 0.141 0.000
AUA3 0.017 -0.064 -0.024 0.057 0.019 0.029 0.087 0.001 0.203 0.081
AUA4 -0.001 0.003 0.015 0.099 -0.032 -0.014 0.052 -0.020 0.089 0.128
      AUA3   AUA4

AS1
AS2
AS3
AS4
AS5
AS6
AS7
AS8
AS9
AF1
AF2
AF3
AF4
MI1
MI2
MI3
MI4
MI5
AUA1
AUA2
AUA3 0.000
AUA4 0.109 0.000

```

```
lavaan::modindices(grmsAAW1fit)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
46	AS1	~~	AS2	12.179	0.097	0.097	0.215	0.215
47	AS1	~~	AS3	3.392	0.048	0.048	0.114	0.114
48	AS1	~~	AS4	2.162	0.037	0.037	0.091	0.091
49	AS1	~~	AS5	7.468	0.072	0.072	0.173	0.173
50	AS1	~~	AS6	0.523	0.016	0.016	0.044	0.044

51	AS1	~~	AS7	0.094	0.008	0.008	0.019	0.019
52	AS1	~~	AS8	0.172	0.011	0.011	0.026	0.026
53	AS1	~~	AS9	0.127	0.008	0.008	0.022	0.022
54	AS1	~~	AF1	1.304	-0.029	-0.029	-0.069	-0.069
55	AS1	~~	AF2	2.360	-0.040	-0.040	-0.096	-0.096
56	AS1	~~	AF3	4.167	-0.053	-0.053	-0.123	-0.123
57	AS1	~~	AF4	10.256	-0.093	-0.093	-0.194	-0.194
58	AS1	~~	MI1	1.050	-0.025	-0.025	-0.062	-0.062
59	AS1	~~	MI2	1.472	-0.033	-0.033	-0.073	-0.073
60	AS1	~~	MI3	2.150	-0.043	-0.043	-0.089	-0.089
61	AS1	~~	MI4	5.074	-0.068	-0.068	-0.136	-0.136
62	AS1	~~	MI5	1.305	0.030	0.030	0.069	0.069
63	AS1	~~	AUA1	0.928	-0.030	-0.030	-0.058	-0.058
64	AS1	~~	AUA2	7.597	-0.073	-0.073	-0.169	-0.169
65	AS1	~~	AUA3	1.581	0.031	0.031	0.076	0.076
66	AS1	~~	AUA4	0.035	-0.005	-0.005	-0.011	-0.011
67	AS2	~~	AS3	2.283	0.044	0.044	0.093	0.093
68	AS2	~~	AS4	9.470	0.085	0.085	0.188	0.188
69	AS2	~~	AS5	0.412	0.019	0.019	0.040	0.040
70	AS2	~~	AS6	7.130	0.066	0.066	0.162	0.162
71	AS2	~~	AS7	3.084	0.049	0.049	0.107	0.107
72	AS2	~~	AS8	3.516	0.054	0.054	0.115	0.115
73	AS2	~~	AS9	10.327	0.078	0.078	0.197	0.197
74	AS2	~~	AF1	6.009	-0.069	-0.069	-0.148	-0.148
75	AS2	~~	AF2	9.086	-0.085	-0.085	-0.186	-0.186
76	AS2	~~	AF3	6.240	-0.071	-0.071	-0.150	-0.150
77	AS2	~~	AF4	1.111	-0.034	-0.034	-0.063	-0.063
78	AS2	~~	MI1	1.749	-0.035	-0.035	-0.079	-0.079
79	AS2	~~	MI2	1.985	-0.042	-0.042	-0.084	-0.084
80	AS2	~~	MI3	3.196	-0.058	-0.058	-0.108	-0.108
81	AS2	~~	MI4	6.072	-0.082	-0.082	-0.147	-0.147
82	AS2	~~	MI5	0.366	-0.018	-0.018	-0.036	-0.036
83	AS2	~~	AUA1	5.910	-0.082	-0.082	-0.146	-0.146
84	AS2	~~	AUA2	1.513	-0.036	-0.036	-0.075	-0.075
85	AS2	~~	AUA3	3.045	-0.047	-0.047	-0.105	-0.105
86	AS2	~~	AUA4	2.871	-0.048	-0.048	-0.103	-0.103
87	AS3	~~	AS4	2.375	0.040	0.040	0.094	0.094
88	AS3	~~	AS5	0.982	-0.027	-0.027	-0.062	-0.062
89	AS3	~~	AS6	0.184	0.010	0.010	0.026	0.026
90	AS3	~~	AS7	1.142	0.028	0.028	0.065	0.065
91	AS3	~~	AS8	0.003	0.001	0.001	0.003	0.003
92	AS3	~~	AS9	0.342	0.013	0.013	0.036	0.036
93	AS3	~~	AF1	0.363	0.016	0.016	0.036	0.036
94	AS3	~~	AF2	0.003	0.002	0.002	0.004	0.004
95	AS3	~~	AF3	0.230	0.013	0.013	0.029	0.029
96	AS3	~~	AF4	0.390	0.019	0.019	0.038	0.038
97	AS3	~~	MI1	2.986	-0.043	-0.043	-0.104	-0.104
98	AS3	~~	MI2	3.984	-0.056	-0.056	-0.120	-0.120

99	AS3	~~	MI3	1.910	-0.042	-0.042	-0.084	-0.084
100	AS3	~~	MI4	0.580	0.024	0.024	0.046	0.046
101	AS3	~~	MI5	1.050	-0.028	-0.028	-0.061	-0.061
102	AS3	~~	AUA1	1.112	-0.034	-0.034	-0.063	-0.063
103	AS3	~~	AUA2	0.326	-0.016	-0.016	-0.035	-0.035
104	AS3	~~	AUA3	0.371	-0.015	-0.015	-0.037	-0.037
105	AS3	~~	AUA4	2.846	-0.045	-0.045	-0.103	-0.103
106	AS4	~~	AS5	0.434	0.017	0.017	0.041	0.041
107	AS4	~~	AS6	2.068	0.032	0.032	0.087	0.087
108	AS4	~~	AS7	0.102	0.008	0.008	0.019	0.019
109	AS4	~~	AS8	0.272	0.014	0.014	0.032	0.032
110	AS4	~~	AS9	6.944	0.058	0.058	0.162	0.162
111	AS4	~~	AF1	0.007	0.002	0.002	0.005	0.005
112	AS4	~~	AF2	0.139	0.010	0.010	0.023	0.023
113	AS4	~~	AF3	1.944	-0.036	-0.036	-0.083	-0.083
114	AS4	~~	AF4	1.621	-0.037	-0.037	-0.076	-0.076
115	AS4	~~	MI1	1.650	-0.031	-0.031	-0.077	-0.077
116	AS4	~~	MI2	2.543	-0.043	-0.043	-0.096	-0.096
117	AS4	~~	MI3	0.803	-0.026	-0.026	-0.054	-0.054
118	AS4	~~	MI4	6.459	-0.076	-0.076	-0.152	-0.152
119	AS4	~~	MI5	0.655	0.021	0.021	0.048	0.048
120	AS4	~~	AUA1	5.187	-0.070	-0.070	-0.136	-0.136
121	AS4	~~	AUA2	1.553	-0.033	-0.033	-0.076	-0.076
122	AS4	~~	AUA3	2.264	-0.036	-0.036	-0.091	-0.091
123	AS4	~~	AUA4	9.746	-0.081	-0.081	-0.190	-0.190
124	AS5	~~	AS6	10.401	0.075	0.075	0.200	0.200
125	AS5	~~	AS7	0.764	-0.023	-0.023	-0.055	-0.055
126	AS5	~~	AS8	0.203	0.012	0.012	0.028	0.028
127	AS5	~~	AS9	7.539	0.063	0.063	0.173	0.173
128	AS5	~~	AF1	1.444	-0.032	-0.032	-0.074	-0.074
129	AS5	~~	AF2	2.768	-0.044	-0.044	-0.105	-0.105
130	AS5	~~	AF3	2.073	-0.039	-0.039	-0.088	-0.088
131	AS5	~~	AF4	0.290	-0.016	-0.016	-0.033	-0.033
132	AS5	~~	MI1	2.913	-0.043	-0.043	-0.105	-0.105
133	AS5	~~	MI2	0.090	-0.008	-0.008	-0.018	-0.018
134	AS5	~~	MI3	0.086	-0.009	-0.009	-0.018	-0.018
135	AS5	~~	MI4	4.037	-0.063	-0.063	-0.123	-0.123
136	AS5	~~	MI5	0.690	0.023	0.023	0.051	0.051
137	AS5	~~	AUA1	3.841	-0.062	-0.062	-0.120	-0.120
138	AS5	~~	AUA2	0.272	-0.014	-0.014	-0.033	-0.033
139	AS5	~~	AUA3	2.430	-0.039	-0.039	-0.096	-0.096
140	AS5	~~	AUA4	0.837	0.025	0.025	0.057	0.057
141	AS6	~~	AS7	2.648	-0.037	-0.037	-0.098	-0.098
142	AS6	~~	AS8	0.025	0.004	0.004	0.010	0.010
143	AS6	~~	AS9	5.033	0.044	0.044	0.136	0.136
144	AS6	~~	AF1	4.871	-0.050	-0.050	-0.132	-0.132
145	AS6	~~	AF2	1.631	-0.029	-0.029	-0.078	-0.078
146	AS6	~~	AF3	0.657	-0.019	-0.019	-0.048	-0.048

147	AS6	~~	AF4	2.212	-0.038	-0.038	-0.089	-0.089
148	AS6	~~	MI1	0.971	-0.021	-0.021	-0.059	-0.059
149	AS6	~~	MI2	0.111	0.008	0.008	0.020	0.020
150	AS6	~~	MI3	0.006	-0.002	-0.002	-0.004	-0.004
151	AS6	~~	MI4	0.525	-0.019	-0.019	-0.043	-0.043
152	AS6	~~	MI5	0.035	-0.004	-0.004	-0.011	-0.011
153	AS6	~~	AUA1	0.100	0.009	0.009	0.019	0.019
154	AS6	~~	AUA2	2.466	-0.037	-0.037	-0.095	-0.095
155	AS6	~~	AUA3	4.354	-0.045	-0.045	-0.125	-0.125
156	AS6	~~	AUA4	2.674	-0.038	-0.038	-0.099	-0.099
157	AS7	~~	AS8	5.121	0.060	0.060	0.138	0.138
158	AS7	~~	AS9	0.072	-0.006	-0.006	-0.016	-0.016
159	AS7	~~	AF1	0.231	-0.012	-0.012	-0.029	-0.029
160	AS7	~~	AF2	2.175	-0.038	-0.038	-0.091	-0.091
161	AS7	~~	AF3	1.070	0.027	0.027	0.062	0.062
162	AS7	~~	AF4	1.097	-0.030	-0.030	-0.063	-0.063
163	AS7	~~	MI1	0.516	-0.017	-0.017	-0.043	-0.043
164	AS7	~~	MI2	0.266	0.014	0.014	0.031	0.031
165	AS7	~~	MI3	0.587	-0.022	-0.022	-0.046	-0.046
166	AS7	~~	MI4	0.910	0.029	0.029	0.057	0.057
167	AS7	~~	MI5	0.450	-0.018	-0.018	-0.040	-0.040
168	AS7	~~	AUA1	0.148	0.012	0.012	0.023	0.023
169	AS7	~~	AUA2	0.249	-0.013	-0.013	-0.030	-0.030
170	AS7	~~	AUA3	0.141	-0.009	-0.009	-0.023	-0.023
171	AS7	~~	AUA4	0.031	0.005	0.005	0.011	0.011
172	AS8	~~	AS9	0.127	0.008	0.008	0.022	0.022
173	AS8	~~	AF1	0.436	0.018	0.018	0.040	0.040
174	AS8	~~	AF2	1.426	-0.032	-0.032	-0.074	-0.074
175	AS8	~~	AF3	0.839	0.025	0.025	0.055	0.055
176	AS8	~~	AF4	3.966	-0.060	-0.060	-0.120	-0.120
177	AS8	~~	MI1	2.597	-0.040	-0.040	-0.097	-0.097
178	AS8	~~	MI2	0.413	0.018	0.018	0.039	0.039
179	AS8	~~	MI3	1.199	0.033	0.033	0.066	0.066
180	AS8	~~	MI4	3.922	-0.062	-0.062	-0.119	-0.119
181	AS8	~~	MI5	0.508	-0.020	-0.020	-0.043	-0.043
182	AS8	~~	AUA1	1.089	-0.033	-0.033	-0.063	-0.063
183	AS8	~~	AUA2	0.538	-0.020	-0.020	-0.045	-0.045
184	AS8	~~	AUA3	3.626	-0.048	-0.048	-0.115	-0.115
185	AS8	~~	AUA4	0.038	-0.005	-0.005	-0.012	-0.012
186	AS9	~~	AF1	3.546	-0.042	-0.042	-0.114	-0.114
187	AS9	~~	AF2	1.941	-0.031	-0.031	-0.086	-0.086
188	AS9	~~	AF3	3.584	-0.043	-0.043	-0.114	-0.114
189	AS9	~~	AF4	0.022	0.004	0.004	0.009	0.009
190	AS9	~~	MI1	9.744	-0.066	-0.066	-0.188	-0.188
191	AS9	~~	MI2	5.385	-0.054	-0.054	-0.140	-0.140
192	AS9	~~	MI3	0.161	-0.010	-0.010	-0.024	-0.024
193	AS9	~~	MI4	0.430	-0.017	-0.017	-0.039	-0.039
194	AS9	~~	MI5	2.516	-0.037	-0.037	-0.095	-0.095

195	AS9	~~	AUA1	7.261	-0.072	-0.072	-0.162	-0.162
196	AS9	~~	AUA2	0.151	-0.009	-0.009	-0.024	-0.024
197	AS9	~~	AUA3	0.422	0.014	0.014	0.039	0.039
198	AS9	~~	AUA4	1.362	-0.027	-0.027	-0.071	-0.071
199	AF1	~~	AF2	17.012	0.107	0.107	0.250	0.250
200	AF1	~~	AF3	31.010	0.146	0.146	0.329	0.329
201	AF1	~~	AF4	20.642	0.134	0.134	0.269	0.269
202	AF1	~~	MI1	2.332	0.037	0.037	0.090	0.090
203	AF1	~~	MI2	1.393	-0.032	-0.032	-0.070	-0.070
204	AF1	~~	MI3	1.235	-0.033	-0.033	-0.066	-0.066
205	AF1	~~	MI4	0.040	-0.006	-0.006	-0.012	-0.012
206	AF1	~~	MI5	0.475	-0.019	-0.019	-0.041	-0.041
207	AF1	~~	AUA1	5.577	-0.073	-0.073	-0.140	-0.140
208	AF1	~~	AUA2	0.079	-0.007	-0.007	-0.017	-0.017
209	AF1	~~	AUA3	1.621	-0.031	-0.031	-0.076	-0.076
210	AF1	~~	AUA4	0.421	0.017	0.017	0.039	0.039
211	AF2	~~	AF3	8.596	0.077	0.077	0.177	0.177
212	AF2	~~	AF4	14.860	0.113	0.113	0.234	0.234
213	AF2	~~	MI1	5.661	0.058	0.058	0.144	0.144
214	AF2	~~	MI2	0.627	-0.022	-0.022	-0.048	-0.048
215	AF2	~~	MI3	0.002	-0.001	-0.001	-0.002	-0.002
216	AF2	~~	MI4	1.339	0.035	0.035	0.070	0.070
217	AF2	~~	MI5	0.058	0.006	0.006	0.015	0.015
218	AF2	~~	AUA1	4.025	0.062	0.062	0.121	0.121
219	AF2	~~	AUA2	0.010	0.003	0.003	0.006	0.006
220	AF2	~~	AUA3	0.175	0.010	0.010	0.025	0.025
221	AF2	~~	AUA4	0.001	-0.001	-0.001	-0.002	-0.002
222	AF3	~~	AF4	5.924	0.072	0.072	0.144	0.144
223	AF3	~~	MI1	0.997	0.025	0.025	0.059	0.059
224	AF3	~~	MI2	0.227	0.013	0.013	0.028	0.028
225	AF3	~~	MI3	1.393	-0.035	-0.035	-0.070	-0.070
226	AF3	~~	MI4	0.467	0.021	0.021	0.040	0.040
227	AF3	~~	MI5	0.026	0.004	0.004	0.010	0.010
228	AF3	~~	AUA1	0.283	-0.017	-0.017	-0.031	-0.031
229	AF3	~~	AUA2	0.149	-0.010	-0.010	-0.023	-0.023
230	AF3	~~	AUA3	1.861	-0.034	-0.034	-0.081	-0.081
231	AF3	~~	AUA4	0.004	0.002	0.002	0.004	0.004
232	AF4	~~	MI1	0.823	0.025	0.025	0.054	0.054
233	AF4	~~	MI2	0.024	-0.005	-0.005	-0.009	-0.009
234	AF4	~~	MI3	0.649	-0.027	-0.027	-0.048	-0.048
235	AF4	~~	MI4	3.299	0.063	0.063	0.107	0.107
236	AF4	~~	MI5	0.300	0.017	0.017	0.032	0.032
237	AF4	~~	AUA1	1.556	-0.044	-0.044	-0.074	-0.074
238	AF4	~~	AUA2	0.081	0.009	0.009	0.017	0.017
239	AF4	~~	AUA3	0.264	-0.014	-0.014	-0.031	-0.031
240	AF4	~~	AUA4	0.122	0.010	0.010	0.021	0.021
241	MI1	~~	MI2	4.445	0.054	0.054	0.124	0.124
242	MI1	~~	MI3	5.139	0.063	0.063	0.134	0.134

243	MI1	~~	MI4	6.788	0.075	0.075	0.153	0.153
244	MI1	~~	MI5	2.059	-0.036	-0.036	-0.084	-0.084
245	MI1	~~	AUA1	3.537	0.055	0.055	0.111	0.111
246	MI1	~~	AUA2	0.727	0.021	0.021	0.051	0.051
247	MI1	~~	AUA3	1.475	0.028	0.028	0.072	0.072
248	MI1	~~	AUA4	4.936	0.055	0.055	0.133	0.133
249	MI2	~~	MI3	18.466	0.134	0.134	0.254	0.254
250	MI2	~~	MI4	2.230	0.048	0.048	0.088	0.088
251	MI2	~~	MI5	4.296	0.059	0.059	0.122	0.122
252	MI2	~~	AUA1	2.548	0.052	0.052	0.094	0.094
253	MI2	~~	AUA2	2.971	0.048	0.048	0.103	0.103
254	MI2	~~	AUA3	0.163	0.010	0.010	0.024	0.024
255	MI2	~~	AUA4	0.514	-0.020	-0.020	-0.043	-0.043
256	MI3	~~	MI4	5.427	0.082	0.082	0.137	0.137
257	MI3	~~	MI5	0.076	-0.009	-0.009	-0.016	-0.016
258	MI3	~~	AUA1	5.439	0.083	0.083	0.138	0.138
259	MI3	~~	AUA2	1.202	0.033	0.033	0.066	0.066
260	MI3	~~	AUA3	0.396	0.018	0.018	0.037	0.037
261	MI3	~~	AUA4	0.099	-0.009	-0.009	-0.019	-0.019
262	MI4	~~	MI5	0.920	0.030	0.030	0.056	0.056
263	MI4	~~	AUA1	0.621	0.029	0.029	0.046	0.046
264	MI4	~~	AUA2	0.166	0.013	0.013	0.024	0.024
265	MI4	~~	AUA3	3.337	0.053	0.053	0.108	0.108
266	MI4	~~	AUA4	1.327	0.036	0.036	0.069	0.069
267	MI5	~~	AUA1	0.485	0.023	0.023	0.041	0.041
268	MI5	~~	AUA2	0.556	0.021	0.021	0.044	0.044
269	MI5	~~	AUA3	0.000	0.000	0.000	0.001	0.001
270	MI5	~~	AUA4	0.201	-0.012	-0.012	-0.027	-0.027
271	AUA1	~~	AUA2	9.802	0.100	0.100	0.187	0.187
272	AUA1	~~	AUA3	18.518	0.127	0.127	0.255	0.255
273	AUA1	~~	AUA4	3.983	0.063	0.063	0.119	0.119
274	AUA2	~~	AUA3	3.471	0.047	0.047	0.112	0.112
275	AUA2	~~	AUA4	9.687	0.084	0.084	0.189	0.189
276	AUA3	~~	AUA4	6.383	0.063	0.063	0.152	0.152

Kline recommends evaluating the “cor” residuals. In our output, these seem to be the “cor.bollen” and are near the bottom. He recommends that residuals  $> .10$  may be possible sources for misfit. He also indicated that patterns may be helpful (is there an item that has consistently high residuals).

Kline also cautions that there is no dependable or trustworthy connection between the size of the residual and the type or degree of model misspecification.

My first read of our results is that the items in the AS# factor were well-defined. I suspect that a multi-factor solution will improve the fit.

The *semTable* package can help us extract the values into a .csv file which will make it easier to create an APA style table. It takes some tinkering...

```
#library(semTable)
#I took out commas internal to the items because the comma causes the text to split across col
v1 <- c(AS1 = "Others expect me to be submissive", AS2 = "Others have been surprised when I dis
grmsAAW1table <- semTable::semTable(grmsAAW1fit, columns = c("eststars", "se", "p"), columnLab
#Can change "print.results" to TRUE if you want to see the (messy) output in the .rmd file (it
```

Cool, but it doesn't contain standardized estimates. One way to get them is to create an updated model with the standardized output:

```
grmsAAW1stdzd <- update (grmsAAW1fit, std.lv = TRUE, std.ov = TRUE, meanstructure = TRUE)
```

Now request both models in the semTable

```
grmsAAW1table <- semTable::semTable(list ("Ordinary" = grmsAAW1fit, "Standardized" = grmsAAW1s
#Can change "print.results" to TRUE if you want to see the (messy) output in the .rmd file (it
```

*Troubleshooting* If, while working with this function you get the error, “Error in file(file, ifelse append, “a”, “w”)) : cannot open the connection” it’s because the .csv file that received your table is still open. R is just trying to write over it. A similar error happens when knitting, or updating any spreadsheet or word document.

### APA Style Results from the Unidimensional model

**Model testing.** To evaluate the models we, we used confirmatory factor analysis (CFA) in the R package, *lavaan* (v.0-6.9) with maximum likelihood estimation. Our sample size was 304. We selected fit criteria for their capacity to assess different aspects of the statistical analysis. As is common among SEM researchers, we reported the Chi-square goodness of fit ( $\chi^2$ ). This evaluates the discrepancy between the unrestricted sample matrix and the restricted covariance matrix. Although the associated *p* value indicates adequate fit when the value is non-significant, it is widely recognized that large sample size can result in a statistically significant *p* value [Byrne, 2016]. The comparative fit index (CFI) is an incremental index, comparing the hypothesized model at least .90 and perhaps higher than .95 [Kline, 2016]. The root mean square error of approximation (RMSEA) takes into account the error of approximation in the population and expresses it per degree of freedom. As such, the fit indicator considers the complexity of the model. Ideal values are equal to or less than .05, values less than .08 represent reasonable fit, and values between .08 and .10 represent mediocre fit. The standardized root mean residual is a standardized measure of the mean absolute covariance residual – the overall difference between the observed and predicted correlations. Values greater than .10 may indicate poor fit and inspection of residuals is then advised. Kline [2016] advised researchers to be cautious when using these criteria as strict cut-offs. Elements such as sample size and model complexity should be considered when evaluating fit.

Our first model was unidimensional where each of the 24 items loaded onto a single factor representing overall, gendered racial microaggressions towards Asian American women. The Chi-square index was statistically significant ( $\chi^2(209) = 1004.136, p < .001$ ) indicating likely misfit. The CFI value of .58 indicated poor fit. The RMSEA = .11 (90% CI [.11, .20]) suggested serious problems. The SRMR value of .12 exceeded the warning criteria of .10. The AIC and BIC values were 17755.028 and 17918.577, respectively, and will become useful in comparing subsequent models.

### 10.5.2 Modeling the GRMSAAW as a First-Order, 4-factor model

#### 10.5.2.1 Specifying and Running the Model

As we know from the article, the GRMSAAW has four subscales. Therefore, let's respecify it as a first-order, four-factor model, allowing the factors to correlate.

**Model identification** is always a consideration. In a multi-dimensional model, each factor requires a minimum of two items/indicators. Our shortest scales are the AF and AUA scales, each with 4 items, so we are OK!

We will be using the `cfa()` function in lavaan. When we do this, it does three things by default:

1. The factor loading of the first indicator of a latent variable is fixed to 1.0; this fixes the scale of the LV
  2. Residual variances are added automatically.
  3. All exogenous LVs are correlated.
- If you are specifying an orthogonal model you will want to switch off the default behavior by including the statement: `auto.cov.lv.x=FALSE`

```
grmsAAW4mod <- 'AS =~ AS1 + AS2 + AS3 + AS4 + AS5 + AS6 + AS7 + AS8 + AS9
                  AF =~ AF1 + AF2 + AF3 + AF4
                  MI =~ MI1 + MI2 + MI3 + MI4 + MI5
                  AUA =~ AUA1 + AUA2 + AUA3 + AUA4'
```

grmsAAW4mod

```
[1] "AS =~ AS1 + AS2 + AS3 + AS4 + AS5 + AS6 + AS7 + AS8 + AS9\n"                                AF =~ AF1 + AF2 +
#This code is identical to the one we ran above -- in this code below, we are just clearly spe


```
grmsAAW4mod <- 'AS =~ AS1 + AS2 + AS3 + AS4 + AS5 + AS6 + AS7 + AS8 + AS9
                  AF =~ AF1 + AF2 + AF3 + AF4
                  MI =~ MI1 + MI2 + MI3 + MI4 + MI5
                  AUA =~ AUA1 + AUA2 + AUA3 + AUA4'
#covariances in our oblique model
AS ~~ AF
AS ~~ MI
AS ~~ AUA
AF ~~ MI
AF ~~ AUA
MI ~~ AUA
```


```

```
grmsAAW4fit <- lavaan::cfa (grmsAAW4mod, data = dfGRMSAAW)
lavaan::summary(grmsAAW4fit, fit.measures=TRUE, standardized=TRUE, rsquare = TRUE)
```

lavaan 0.6.17 ended normally after 61 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	50
Number of observations	304

#### Model Test User Model:

Test statistic	251.925
Degrees of freedom	203
P-value (Chi-square)	0.011

#### Model Test Baseline Model:

Test statistic	1479.910
Degrees of freedom	231
P-value	0.000

#### User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.961
Tucker-Lewis Index (TLI)	0.955

#### Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-6999.298
Loglikelihood unrestricted model (H1)	-6873.336
Akaike (AIC)	14098.596
Bayesian (BIC)	14284.448
Sample-size adjusted Bayesian (SABIC)	14125.873

#### Root Mean Square Error of Approximation:

RMSEA	0.028
90 Percent confidence interval - lower	0.014
90 Percent confidence interval - upper	0.039
P-value H_0: RMSEA <= 0.050	1.000
P-value H_0: RMSEA >= 0.080	0.000

#### Standardized Root Mean Square Residual:

SRMR	0.048
------	-------

#### Parameter Estimates:

	Standard errors	Standard				
	Information	Expected				
	Information saturated (h1) model	Structured				
<b>Latent Variables:</b>						
	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
AS =~						
AS1	1.000				0.458	0.600
AS2	1.080	0.131	8.242	0.000	0.495	0.604
AS3	0.915	0.121	7.556	0.000	0.419	0.538
AS4	0.905	0.116	7.792	0.000	0.415	0.560
AS5	1.142	0.133	8.579	0.000	0.523	0.639
AS6	0.709	0.099	7.174	0.000	0.325	0.504
AS7	0.808	0.114	7.104	0.000	0.370	0.498
AS8	0.913	0.121	7.536	0.000	0.418	0.536
AS9	0.847	0.105	8.098	0.000	0.388	0.589
AF =~						
AF1	1.000				0.444	0.614
AF2	1.201	0.151	7.980	0.000	0.534	0.691
AF3	0.836	0.124	6.736	0.000	0.371	0.516
AF4	1.010	0.143	7.049	0.000	0.449	0.550
MI =~						
MI1	1.000				0.330	0.492
MI2	1.147	0.202	5.669	0.000	0.379	0.506
MI3	1.369	0.232	5.890	0.000	0.452	0.543
MI4	1.118	0.213	5.252	0.000	0.369	0.445
MI5	0.765	0.174	4.388	0.000	0.253	0.344
AUA =~						
AUA1	1.000				0.434	0.509
AUA2	1.061	0.162	6.562	0.000	0.460	0.601
AUA3	0.864	0.139	6.222	0.000	0.375	0.543
AUA4	1.037	0.159	6.503	0.000	0.449	0.590
<b>Covariances:</b>						
	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
AS ~~						
AF	0.110	0.021	5.257	0.000	0.541	0.541
MI	0.083	0.018	4.727	0.000	0.551	0.551
AUA	0.117	0.023	5.054	0.000	0.589	0.589
AF ~~						
MI	0.087	0.018	4.704	0.000	0.590	0.590
AUA	0.106	0.023	4.713	0.000	0.553	0.553
MI ~~						
AUA	0.109	0.022	4.878	0.000	0.760	0.760
<b>Variances:</b>						
	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.AS1	0.372	0.034	10.872	0.000	0.372	0.639

.AS2	0.427	0.039	10.847	0.000	0.427	0.635
.AS3	0.431	0.038	11.279	0.000	0.431	0.711
.AS4	0.377	0.034	11.151	0.000	0.377	0.686
.AS5	0.397	0.038	10.549	0.000	0.397	0.592
.AS6	0.310	0.027	11.452	0.000	0.310	0.746
.AS7	0.416	0.036	11.481	0.000	0.416	0.752
.AS8	0.433	0.038	11.289	0.000	0.433	0.712
.AS9	0.283	0.026	10.955	0.000	0.283	0.653
.AF1	0.326	0.034	9.615	0.000	0.326	0.623
.AF2	0.311	0.038	8.230	0.000	0.311	0.522
.AF3	0.379	0.035	10.724	0.000	0.379	0.734
.AF4	0.464	0.045	10.402	0.000	0.464	0.697
.MI1	0.341	0.032	10.648	0.000	0.341	0.758
.MI2	0.417	0.040	10.517	0.000	0.417	0.744
.MI3	0.487	0.048	10.099	0.000	0.487	0.705
.MI4	0.551	0.050	11.041	0.000	0.551	0.802
.MI5	0.476	0.041	11.639	0.000	0.476	0.882
.AUA1	0.538	0.050	10.749	0.000	0.538	0.741
.AUA2	0.374	0.038	9.730	0.000	0.374	0.639
.AUA3	0.337	0.032	10.431	0.000	0.337	0.705
.AUA4	0.378	0.038	9.882	0.000	0.378	0.652
AS	0.210	0.040	5.283	0.000	1.000	1.000
AF	0.197	0.039	5.006	0.000	1.000	1.000
MI	0.109	0.029	3.789	0.000	1.000	1.000
AUA	0.188	0.047	4.040	0.000	1.000	1.000

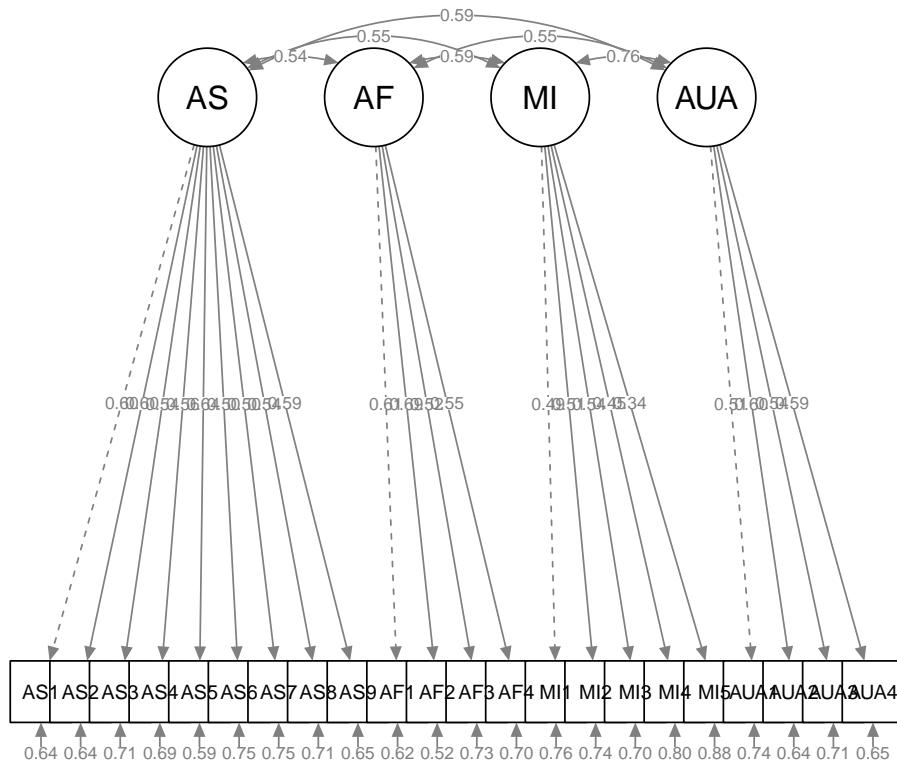
R-Square:

## Estimate

AS1	0.361
AS2	0.365
AS3	0.289
AS4	0.314
AS5	0.408
AS6	0.254
AS7	0.248
AS8	0.288
AS9	0.347
AF1	0.377
AF2	0.478
AF3	0.266
AF4	0.303
MI1	0.242
MI2	0.256
MI3	0.295
MI4	0.198
MI5	0.118
AUA1	0.259
AUA2	0.361

AUA3	0.295
AUA4	0.348

```
semPlot::semPaths(grmsAAW4fit, layout = "tree", style = "lisrel", what = "col", whatLabels = "none")
```



The table

First an update to get the standardized results:

```
grmsAAW4stdzd <- update (grmsAAW4fit, std.lv = TRUE, std.ov = TRUE, meanstructure = TRUE)
```

```
grmsAAW4table <- semTable::semTable(list ("Ordinary" = grmsAAW4fit, "Standardized" = grmsAAW4stdzd))
```

#Can change "print.results" to TRUE if you want to see the (messy) output in the .rmd file (it

### 10.5.2.2 Interpretation

Our model converged, normally, with 37 iterations. The estimator was the lavaan default, maximum likelihood (ML). All 304 cases were used in the analysis.

I mapped our pattern coefficients into the GRMSAAW tables. Most pattern coefficients are strong, significant, and stably connected to their respective factor. The lowest factor loading was .220 (MI5).

A multidimensional factor structure also includes correlations/covariances between factors. We can see that the correlation (look at the Std.all column) shows the following correlations (none are statistically significant):

AF & AS: 0.017 AF & MI: -0.060 AF & AUA: 0.035 AS & MI: 0.082 AS & AUA: 0.035 MI & AUA: 0.077

For our multi-dimensional GRMSAAW4 CFA  $\chi^2(203) = 220.858, p < .186$ , this significant value is not what we want because it says that our specified model is not statistically significantly different than the covariances in the model. That is, our more parsimonious model is a reasonable explanation (simplification).

The CFI and TLI compare user (the 4-dimensional model we specified) and baseline (where no relations would exist between variables) models. These values will always be close together because the only difference is that the TLI imposes a penalty for any model complexity. The CFI seems to be more commonly reported and its value is 0.991. This means our model performed 99% better than a model with no relations. It well-exceeds the traditional cutoffs of .90 and the more strict cutoff of .95. The TLI imposes a greater relative penalty for model complexity, consequently it is a smidge lower at .989.

The RMSEA one of the *badness of fit*, absolute fit index, statistics where a value of 0.00 is the best fit. Our RMSEA = 0.017 (90%CI[.000, .031]). As a quick reminder, an there is general consensus that  $RMSEA \leq .05$  is desired and an  $RMSEA \geq .10$  points to serious problems. We watch the upper bound of the confidence interval to see that it isn't sneaking into the danger zone.

The SRMR is another absolute, *badness of fit* index (i.e., perfect model fit is when the value = 0.00 and increasingly higher values indicate the “badness”). The SRMR is a measure of the mean absolute covariance residual. Standardizing the value facilitates interpretation. Poor fit is indicated when  $SRMR \geq .10$ . The GRMSAAW SRMR = .058.

Recall, Hu and Bentler's **combination rule** (which is somewhat contested) suggested that the SRMR be interpreted along with the CFI such that:  $CFI \geq .95$  and  $SRMR \leq .08$ .

For our unidimensional GRMSAAW CFA, the CFI = .99 and the SRMR = .058. We are close!

The AIC and BIC utilize an information theory approach to data analysis by combining statistical estimation and model selection into a single framework. The BIC augments the AIC by taking sample size into consideration. We can compare the values from our current model to the former one. The model with the smallest value of the predictive fit index is chosen as the one that is most likely to replicate. It means that this model has relatively better fit and fewer free parameters than competing models. We will do that in the next section.

Before moving to model comparison, it is a good practice for locating sources of misfit (we look for relatively large values) is to inspect the residuals, so let's do that.

```
lavaan::fitted(grmsAAW4fit)
```

```
$cov
      AS1   AS2   AS3   AS4   AS5   AS6   AS7   AS8   AS9   AF1   AF2   AF3
AS1  0.582
AS2  0.227  0.672
AS3  0.192  0.207  0.607
AS4  0.190  0.205  0.174  0.548
AS5  0.240  0.259  0.219  0.217  0.671
AS6  0.149  0.161  0.136  0.135  0.170  0.416
AS7  0.170  0.183  0.155  0.154  0.194  0.120  0.553
AS8  0.192  0.207  0.175  0.173  0.219  0.136  0.155  0.608
AS9  0.178  0.192  0.163  0.161  0.203  0.126  0.144  0.162  0.434
AF1  0.110  0.119  0.101  0.100  0.126  0.078  0.089  0.100  0.093  0.524
AF2  0.132  0.143  0.121  0.120  0.151  0.094  0.107  0.121  0.112  0.237  0.596
AF3  0.092  0.099  0.084  0.083  0.105  0.065  0.074  0.084  0.078  0.165  0.198  0.517
AF4  0.111  0.120  0.102  0.101  0.127  0.079  0.090  0.101  0.094  0.199  0.239  0.167
MI1  0.083  0.090  0.076  0.075  0.095  0.059  0.067  0.076  0.071  0.087  0.104  0.072
MI2  0.096  0.103  0.087  0.087  0.109  0.068  0.077  0.087  0.081  0.099  0.119  0.083
MI3  0.114  0.123  0.104  0.103  0.130  0.081  0.092  0.104  0.097  0.118  0.142  0.099
MI4  0.093  0.101  0.085  0.084  0.106  0.066  0.075  0.085  0.079  0.097  0.116  0.081
MI5  0.064  0.069  0.058  0.058  0.073  0.045  0.052  0.058  0.054  0.066  0.080  0.055
AUA1 0.117  0.126  0.107  0.106  0.133  0.083  0.094  0.107  0.099  0.106  0.128  0.089
AUA2 0.124  0.134  0.113  0.112  0.142  0.088  0.100  0.113  0.105  0.113  0.136  0.094
AUA3 0.101  0.109  0.092  0.091  0.115  0.072  0.082  0.092  0.086  0.092  0.111  0.077
AUA4 0.121  0.131  0.111  0.110  0.138  0.086  0.098  0.111  0.103  0.110  0.133  0.092
      AF4   MI1   MI2   MI3   MI4   MI5   AUA1   AUA2   AUA3   AUA4
AS1
AS2
AS3
AS4
AS5
AS6
AS7
AS8
AS9
AF1
AF2
AF3
AF4  0.665
MI1  0.087  0.450
MI2  0.100  0.125  0.561
MI3  0.120  0.149  0.171  0.691
MI4  0.098  0.122  0.140  0.167  0.688
MI5  0.067  0.083  0.096  0.114  0.093  0.540
AUA1 0.108  0.109  0.125  0.149  0.122  0.083  0.726
AUA2 0.114  0.115  0.132  0.158  0.129  0.088  0.199  0.585
```

```
AUA3 0.093 0.094 0.108 0.129 0.105 0.072 0.163 0.172 0.477
AUA4 0.111 0.113 0.129 0.154 0.126 0.086 0.195 0.207 0.168 0.580
```

```
#lavaan::residuals(grmsAAW4fit, type = "raw")
#lavaan::residuals(grmsAAW4fit, type = "standardized")
lavaan::residuals(grmsAAW4fit, type = "cor")
```

```
$type
[1] "cor.bollen"

$cov
      AS1     AS2     AS3     AS4     AS5     AS6     AS7     AS8     AS9     AF1
AS1  0.000
AS2  0.047  0.000
AS3  0.030 -0.003  0.000
AS4 -0.006  0.038  0.020  0.000
AS5  0.042 -0.056 -0.063 -0.029  0.000
AS6 -0.027  0.035 -0.019  0.005  0.074  0.000
AS7 -0.015  0.022  0.034 -0.016 -0.046 -0.094  0.000
AS8 -0.024  0.012 -0.017 -0.020 -0.011 -0.029  0.082  0.000
AS9 -0.053  0.036 -0.020  0.038  0.041  0.033 -0.041 -0.028  0.000
AF1 -0.024 -0.100  0.060  0.023 -0.010 -0.081  0.018  0.062 -0.059  0.000
AF2  0.016 -0.066  0.085  0.079  0.036  0.007  0.029  0.038  0.016 -0.030
AF3 -0.059 -0.098  0.058 -0.038 -0.017 -0.016  0.086  0.077 -0.057  0.094
AF4 -0.100 -0.028  0.070 -0.026  0.027 -0.040  0.001 -0.042  0.038  0.034
MI1 -0.007 -0.038 -0.031 -0.026 -0.019 -0.017  0.015 -0.026 -0.102  0.043
MI2 -0.019 -0.046 -0.046 -0.043  0.033  0.041  0.067  0.069 -0.071 -0.089
MI3 -0.014 -0.048 -0.004  0.002  0.050  0.033  0.025  0.101  0.026 -0.078
MI4 -0.069 -0.096  0.070 -0.090 -0.041 -0.011  0.083 -0.051 -0.002 -0.038
MI5  0.121  0.031  0.030  0.101  0.120  0.050  0.046  0.044 -0.001 -0.018
AUA1 -0.028 -0.110 -0.023 -0.092 -0.054  0.024  0.045 -0.023 -0.107 -0.140
AUA2 -0.058 -0.021  0.038 -0.007  0.049 -0.027  0.044  0.031  0.031 -0.019
AUA3  0.090 -0.055  0.022 -0.032 -0.007 -0.063  0.036 -0.032  0.062 -0.069
AUA4  0.053 -0.030  0.002 -0.074  0.111 -0.021  0.080  0.062  0.009  0.030
          AF2     AF3     AF4     MI1     MI2     MI3     MI4     MI5    AUA1    AUA2
AS1
AS2
AS3
AS4
AS5
AS6
AS7
AS8
AS9
AF1
AF2  0.000
AF3 -0.035  0.000
```

```

AF4 -0.005 -0.027  0.000
MI1  0.100  0.030  0.027  0.000
MI2  -0.039  0.001 -0.028 -0.013  0.000
MI3  0.004 -0.072 -0.051 -0.009  0.080  0.000
MI4  0.046  0.015  0.071  0.028 -0.033  0.005  0.000
MI5  0.057  0.026  0.049 -0.115  0.053 -0.058  0.006  0.000
AUA1 0.085 -0.044 -0.076  0.031  0.012  0.047 -0.018  0.022  0.000
AUA2 0.038 -0.015  0.019 -0.010  0.024  0.000 -0.028  0.042  0.006  0.000
AUA3 0.041 -0.067 -0.023  0.008 -0.035 -0.020  0.041  0.003  0.076 -0.041
AUA4 0.043  0.012  0.029  0.059 -0.078 -0.053  0.013 -0.006 -0.036  0.012
      AUA3   AUA4

AS1
AS2
AS3
AS4
AS5
AS6
AS7
AS8
AS9
AF1
AF2
AF3
AF4
MI1
MI2
MI3
MI4
MI5
AUA1
AUA2
AUA3  0.000
AUA4 -0.003  0.000

```

```
#lavaan::modindices(grmsAAW4fit)
```

## 10.6 Model Comparison

We evaluated two models (i.e., a unidimensional model and four-factor correlated model), which one is better? While, we have the narrative comparison (and would create a table with the comparisons) where the four-dimensional fit values (CFI = 0.99, RMSEA = 0.02 (90%CI[.00, .03], and SRMR = .058) outperformed the unidimensional ones (CFI = 0.58, RMSEA = .11 (90%CI[.11, .20]), and SRMR = .12). We can formally compare them with statistical comparisons.

Easy are AIC and BIC comparisons where “smaller value wins.”

AIC GRMSAAW1: 17755.028 AIC GRMSAAW4: 16983.750

BIC GRMSAAW1: 17918.577 BIC GRMSAAW4: 17169.602

In both cases, the smaller values are for the more complex, 4-dimensional model. The interpretation is that the model with the smaller AIC/BIC values is most likely to replicate.

Additionally, the **chi-square difference test**,  $\chi_D^2$  can be used to compare nested models. Single-factor CFA models are nested under any other CFA model with two or more factors *for the same indicators*. This is because a one-factor model is a restricted version of any model with multiple factors. Our unidimensional GRMSAAW was nested under the 4-factor GRMSAAW model.

To calculate the chi-square difference test, we first grab the chi-square test values:

GRMSAAW1:  $\chi^2(209) = 1004.136, p < .001$  GRMSAAW4:  $\chi^2(203) = 220.858, p < .186$

Given both sets of results we calculate:  $\chi^2(6) = 783.278, p < .05$  and determine that the two models are statistically significantly different. Given that the fit statistics are better for the single-order, correlated, four-factor model, we prefer that one.

How did I do that?

- Subtract the df
- Subtract the chi-square values
- Use a chi-square difference table to look up the chi-square critical value for a 6 df test
  - <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm>, or
  - use this code to look it up `qchisq(p, df, lower.tail=FALSE)`
  - the critical value for our test is 12.592
- We conclude that the two models are statistically significantly different; our 4-factor model is preferred.

```
209-203 #subtract df
```

```
[1] 6
```

```
1004.136 - 220.858 #subtract chi-square values
```

```
[1] 783.278
```

```
qchisq(.05, 6, lower.tail=FALSE)
```

```
[1] 12.59159
```

Of course, there is a function for something this easy:

```
lavaan::lavTestLRT(grmsAAW1fit, grmsAAW4fit)
```

### Chi-Squared Difference Test

```
Df      AIC     BIC   Chisq Chisq diff    RMSEA Df diff
grmsAAW4fit 203 14099 14284 251.92
grmsAAW1fit 209 14301 14465 466.67      214.75 0.33829       6
Pr(>Chisq)
grmsAAW4fit
grmsAAW1fit < 0.0000000000000022 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And we get the same result:  $\chi^2(6) = 783.28, p < .001$

And now a table with estimates and fit indices from both models.

```
#All the requested data gets transferred over, but the pattern coefficients do not end up side by side
#Can change "print.results" to TRUE if you want to see the (messy) output in the .rmd file (it will be messy)
```

`grmsAAWtables <- semTable::semTable(list("Single Dimension" = grmsAAW1fit, "Multidimensional" = grmsAAW4fit))`

Let's try it with standardized output:

```
GRMSAAWstdzd <- semTable::semTable(list("Single Dimension" = grmsAAW1stdzd, "Multidimensional" = grmsAAW4stdzd))
#Can change "print.results" to TRUE if you want to see the (messy) output in the .rmd file (it will be messy)
```

### (Placeholder, more to come!) APA Results Section:

**Model testing.** To evaluate the models we, we used confirmatory factor analysis (CFA) in the R package, *lavaan* (v.0.6-9) with maximum likelihood estimation. Our sample size was 304. We selected fit criteria for their capacity to assess different aspects of the statistical analysis. As is common among SEM researchers, we reported the Chi-square goodness of fit ( $\chi^2$ ). This evaluates the discrepancy between the unrestricted sample matrix and the restricted covariance matrix. Although the associated  $p$  value indicates adequate fit when the value is non-significant, it is widely recognized that large sample size can result in a statistically significant  $p$  value [Byrne, 2016]. The comparative fit index (CFI) is an incremental index, comparing the hypothesized model at least .90 and perhaps higher than .95 [Kline, 2016]. The root mean square error of approximation (RMSEA) takes into account the error of approximation in the population and expresses it per degree of freedom. As such, the fit indicator considers the complexity of the model. Ideal values are equal to or less than .05, values less than .08 represent reasonable fit, and values between .08 and .10 represent mediocre fit. The standardized root mean residual is a standardized measure of the mean absolute covariance residual – the overall difference between the observed and predicted correlations. Values greater than .10 may indicate poor fit and inspection of residuals is then advised. Because we were interested in comparing nested models we used the Chi-square difference test where a significant chi-square indicates statistically significant differences in models. Additionally we used Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) that take model complexity and sample size into consideration. Models with lower values on each are

considered to be superior. Kline (2016) advised researchers to be cautious when using these criteria as strict cut-offs. Elements such as sample size and model complexity should be considered when evaluating fit. Table 1 provides a side-by-side comparison of the resulting parameter estimates and fit statistics; Figures 1 and 2 provide a graphic representation of the models tested.

Our first model was unidimensional where each of the 22 items loaded onto a single factor representing overall gendered racial microaggressions for Asian American women. Standardized pattern coefficients ranged between -.030 and .799 and were not all statistically significant. The Chi-square index was statistically significant ( $\chi^2(209) = 1004.136, p < .001$ ) indicating likely misfit. The CFI value of .58 indicated poor fit. The RMSEA = .11 (90% CI [.11, .20]) suggested serious problems. The SRMR value of .12 exceeded the warning criteria of .10. The AIC and BIC values were 17755.028 and 17918.577, respectively, and will become useful in comparing subsequent models.

Our second model was a single-order, multidimensional model where each of the 22 items loaded onto one of four factors. Standardized pattern coefficients ranged between .59 and .80 on the AF factor, between .64 and .82 on the AS factor, between .35 and .60 on the MI factor, and between .59 and .82 on the AUA factor. The Chi-square index was statistically significant ( $\chi^2(203) = 220.858, p < .186$ ) indicating reasonable fit. The CFI value of .99 exceeded the recommendation of .95. The RMSEA = .017 (90% CI [.000, .031]) was satisfactory. The SRMR value of .058 remained below the warning criteria of .10. The AIC and BIC values were 16983.750 and 17169.602, respectively.

The Chi-square difference test ( $\chi^2(6) = 783.28, p < .001$ ) was statistically significant and AIC and BIC values of the multidimensional value were lowest. Thus, we conclude the multidimensional model (i.e., the first-order, correlated factors model) is superior and acceptable for use in preliminary research and evaluation.

*We will continue to create, evaluate, and compare models in the next lesson.*

## 10.7 A concluding thought

Much like the children's game *Don't Break the Ice* we start with a full, saturated, matrix of sample data where every indicator/item is allowed to correlate/covary with every other.

As researchers, we specify a more parsimonious model where we fix some relations to zero and allow others to relate. In our GRMSAAW example, we allowed

- the AF items to relate via their relationship to the AF factor;
- the AS items to relate via their relationship to the AS factor;
- the MI items to relate via their relationship to the MI factor; and
- the AUA items to relate via their relationship to the AUA factor.
- we did not allow any of the items on any given factor to relate to the items on any other factor; these are *hard hypotheses* where we fix the relation to zero.

Our goal (especially via the chi-square test) is that we account for as much variance as possible through the specified relations that remain. Harkening to the *Don't Break the Ice* metaphor, we want the ice matrix to remain stable with as many ice cubes deleted as possible.



Source: <https://www.flickr.com/photos/arfsb/4407495674>

## 10.8 Practice Problems

In each of these lessons I provide suggestions for practice that allow you to select one or more problems that are graded in difficulty. The least complex is to change the random seed in the research and rework the problem demonstrated in the lesson. The most complex is to use data of your own. In either case, please plan to:

### 10.8.1 Problem #1: Play around with this simulation.

Copy the script for the simulation and then change (at least) one thing in the simulation to see how it impacts the results.

Using the lecture and workflow (chart) as a guide, please work through all the steps listed in the proposed assignment/grading rubric.

Assignment Component	Points Possible	Points Earned
1. Prepare data for CFA (items only df, reverse-scored)	5	_____
2. Specify and run a unidimensional model	5	_____
3. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____

Assignment Component	Points Possible	Points Earned
4. Specify and run a single-order model with correlated factors	5	_____
5. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____
6. Compare model fit with $\chi^2\Delta$ , AIC, BIC	5	_____
7. APA style results with table(s) and figure	5	_____
8. Explanation to grader	5	_____
<b>Totals</b>	<b>40</b>	_____

### 10.8.2 Problem #2: Use simulated data from other lessons.

The second option comes from the “the back of the book” where a **chapter** contains simulated data for all of the examples worked in this volume. Any of these is available for CFA.

Assignment Component	Points Possible	Points Earned
1. Prepare data for CFA (items only df, reverse-scored)	5	_____
2. Specify and run a unidimensional model	5	_____
3. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____
4. Specify and run a single-order model with correlated factors	5	_____
5. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____
6. Compare model fit with $\chi^2\Delta$ , AIC, BIC	5	_____
7. APA style results with table(s) and figure	5	_____
8. Explanation to grader	5	_____
<b>Totals</b>	<b>40</b>	_____

### 10.8.3 Problem #3: Try something entirely new.

As a third option, you are welcome to use data to which you have access and is suitable for CFA. These could include other simulated data, data available through open science repositories, or your own data (presuming you have permission to use it). In either case, please plan to:

Using the lecture and workflow (chart) as a guide, please work through all the steps listed in the proposed assignment/grading rubric.

Assignment Component	Points Possible	Points Earned
1. Prepare data for CFA (items only df, reverse-scored)	5	_____
2. Specify and run a unidimensional model	5	_____
3. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____

Assignment Component	Points Possible	Points Earned
4. Specify and run a single-order model with correlated factors	5	_____
5. Narrate adequacy of fit with $\chi^2$ , CFI, RMSEA, SRMR (write a mini-results section)	5	_____
6. Compare model fit with $\chi^2\Delta$ , AIC, BIC	5	_____
7. APA style results with table(s) and figure	5	_____
8. Explanation to grader	5	_____
<b>Totals</b>	<b>40</b>	_____

# Chapter 11

## Additional Simulations

In each of the suggestions for practice (i.e., “homework”), I suggest that an option for a middle-level degree of opportunity is to use data from another simulation. One option is use simulations from other lessons. However, in this chapter, I provide three simulations that are not used in any of the lessons. Because these are not used in any of the lessons, it is possible to select one simulation and use it for practice in all (i.e., excepting invariance testing, unless otherwise noted) of the lessons. There are a few points to keep in mind:

- With each of these, it would be possible to focus on *any* of the instruments used within the simulation (i.e., not just the focal instrument of the evaluation).
  - There is a caveat, there are times when I cannot locate the original factor loadings and therefore assign a uniform “0.8.”
  - Additionally, I recommend that you select scales with at least three subscales.
- I recommend that you retrieve the original psychometric article so that you can compare your analytic choices and results with those completed in the article.
- I have added the code, “eval = FALSE” in each of the chunks. This prevents the chunk from executing when the OER is *building*. If you copy the script you will need to remove that code.

### 11.1 iBelong Scale

The iBelong Scale [Lee and Neville, 2024] is a measure of racial-ethnic-cultural (REC) belonging for BIPOC people. The scale includes 25 items with responses rated on a 6 point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). At the start of the survey, participants are asked to self-identify their REC identity (i.e., “In terms of my racial-ethnic-cultural [REC] identity I consider myself to be: \_\_\_\_\_”). Higher scores indicate a stronger sense of REC belonging. Lee and Neville [2024] reported that a five-factor solution resulted in the best fit.

Beneath each of the five factors are a list of the items; variable names follow in parentheses.

- Authenticity:
  - In general, I do not have to change my behaviors when I am with REC group members (Auth1)

- I am free to be myself with members of my REC group (Auth2)
- In general, I do not have to hide parts of who I am with members of my REC group (Auth3)
- Members of my REC group see me for who I am (Auth4)
- I feel accepted by my REC group for who I am (Auth5)
- Connection:
  - I have similar experiences to others in my REC group (Conn1)
  - Others in my REC group see the world in a similar way as me (Conn2)
  - My success is connected to the success of members of my REC group (Conn3)
  - I have many things in common with people from my REC group (Conn4)
  - I feel connected to my REC group because of our shared experiences (Conn5)
- Home
  - I feel close to other members of my REC group (Home1)
  - I have shared interests with others in my REC group (Home2)
  - I have shared values with others in my REC group (Home3)
  - In general, I am at ease when I am with people from my REC group (Home4)
  - In general, I feel “at home” when I am with members of my REC group (Home5)
- REC Thriving
  - I am happy that I am a member of my REC group (Thriv1)
  - I am proud to be a member of my REC group (Thriv2)
  - I like learning about my REC group’s history (Thriv3)
  - I value my REC group (Thriv4)
  - I consider it an honor to be a member of my REC group (Thriv5) \*Self-Definition
  - I define what belonging to my REC group means to me (Self1)
  - I am a member of my REC group on my own terms (Self2)
  - I decide what my REC group membership is (Self3)
  - I do not have to prove that I belong to my REC group to others in my REC group (Self4)
  - Others do not define my belonging to my REC group (Self5)

A preprint of the article is available at ResearchGate. Below is the script for simulating item-level data from the factor loadings, means, and sample size presented in the published article.

Six additional scales were reported in the Lee and Neville [2024] article. Unfortunately, I could not locate factor loadings for all of them; in these cases I uniformly assigned the factor loading of 0.8. The scales, their original citation, and information about how I simulated data for each are listed below.

- **General Belongingness Scale** [GBS; Malone et al. [2012]] is a 12-item item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). There are two six-item subscales: Acceptance/Inclusion (“When I am with other people, I feel included”) and Rejection/Exclusion (“I feel like an outsider”). The items on the Rejection/Exclusion subscale must be rescored such that higher scores represent a greater sense of belonging in general. In the iBELONG study [Lee and Neville, 2024], only a total GBS scale was used. Our simulation used factor loadings from Malone et al. [2012] but treated them as a single scale. This means that there is need to reverse-score the Rejection/Exclusion items in this simulated data.

- **The Collective Self-Esteem Scale—Race Specific Version** [CSESR; [Crocker et al. \[1994\]](#)] is a 16-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores represent a stronger sense of membership with one's own racial/ethnic group. There are four, four-item subscales: Membership CSE, Public CSE, Private CSE, and Importance to Identity, however the iBelong study [[Lee and Neville, 2024](#)] used a total scale score. An example item is, “The racial-ethnic group I belong to is an important reflection of who I am.” Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items on the general factor.
- **The Multigroup Ethnic Identity Measure – Revised** [MEIMR; [Brown et al. \[2014\]](#)] is a 7-item scale. The first item is an open-ended question for identification of respondent ethnic group. The remaining six items are assessed with Likert scaling ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). There are two, three-item subscales which assess exploration (e.g., “I have spent time trying to find out more about my ethnic group, such as its history, tradition, and customs”) and commitment (“I have a strong sense of belonging to my own ethnic group.”) Higher scores indicate more positive ethnic identity. In the iBELONG study [[Lee and Neville, 2024](#)], only a total MEIMR scale was used. Our simulation used factor loadings from Brown et al. [[2014](#)] but treated them as a single scale.

My simulation code would not converge when I included other scales. Although I have included code for simulating them, I have hashtags them out. If you were interested in playing with any of these variables you might be able to hashtag out earlier scales and replace them with these (i.e., hashtag out the GBS, CSESR, and MEIMR and include the CHS, ULS, MHI, or SWLS). Note that the “swapping of hashtags” needs to be all places in the code where they occur (i.e., measurement model, means, correlations):

- **The Cultural Homelessness Scale** [CHS; NEED CITATION] is a 14-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is, “I feel that I don't belong to any ethnic or cultural group.” Higher scores indicate greater racial-ethnic cultural homelessness. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items on the general factor.
- **UCLA Loneliness Scale** [ULS8; [Hays and DiMatteo \[1987\]](#)] is an 8-item scale with Likert scaling ranging from 1 (*never*) to 4 (*often*). An example item is, “I lack companionship.” Higher scores indicate higher levels of loneliness. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items on the general factor.
- **Mental Health Inventory** [MHI; [Veit and Ware \[1983\]](#)] is an five-item scale with Likert scaling ranging from 1 (*all of the time*) to 4 (*none of the time*). An example item is, “In the last month, how often have you felt down and blue?” Higher scores indicate higher levels of well-being. Because I was not able to locate factor loadings from a psychometric evaluation, I simulated the data by specifying a 0.8 as a standardized factor loading for each of the items on the general factor.
- **Satisfaction with Life Scale** [SWLS; [Diener et al. \[1985\]](#)] is an five-item scale with Likert scaling ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). An example item is, “In most ways, my life is close to ideal.” Higher scores indicate greater life satisfaction. I simulated data by estimating factor loadings from the source article [[Diener et al., 1985](#)].

Because data is collected at the item level (and I want this resource to be as practical as possible, I have simulated the data for each of the scales at the item level. Simulating the data involved using factor loadings, means, and correlations between the scales. Because the simulation will produce “out-of-bounds” values, the code below re-scales the scores into the range of the Likert-type scaling and rounds them to whole values.

*#Entering the intercorrelations, means, and standard deviations from the journal article*

```
iBelong_generating_model <- '
#measurement model
Authenticity =~ .66*Auth1 + .64*Auth2 + .63*Auth3 + .58*Auth4 + .55*Auth5
Connection =~ .72*Conn1 + .60*Conn2 + .59*Conn3 + .58*Conn4 + .58*Conn5
Home =~ .64*Home1 + .63*Home2 + .57*Home3 + .57*Home4 + .53*Home5
RECthriving =~ .68*Thriv1 + .66*Thriv2 + .59*Thriv3 + .54*Thriv4 + .49*Thriv5
SelfDefinition =~ .82*Self1 + .65*Self2 + .65*Self3 + .54*Self4 + .49*Self5
GBS =~ .70*GBS1 + .67*GBS2 + .65*GBS5 + .67*GBS8 + .70*GBS10 + .78*GBS11 + .78*GBS3 +
CSESR =~ .8*CSESR1 + .8*CSESR2 + .8*CSESR3 + .8*CSESR4 + .8*CSESR5 + .8*CSESR6 + .8*CSESR7
MEIMR =~ .60*MEIMR1 + .93*MEIMR4 + .81*MEIMR5 + .86*MEIMR2 + .84*MEIMR3 + .89*MEIMR6
#CHS =~ .80*CHS1 + .80*CHS2 + .80*CHS3 + .80*CHS4 + .80*CHS5 + .80*CHS6 + .80*CHS7 + .80*CHS8
#ULS =~ .80*ULS1 + .80*ULS2 + .80*ULS3 + .80*ULS4 + .80*ULS5 + .80*ULS6 + .80*ULS7 + .80*ULS8
#MHI =~ .80*MHI1 + .80*MHI2 + .80*MHI3 + .80*MHI4 + .80*MHI5
#SWLS =~ .84*SWLS1 + .77*SWLS2 + .83*SWLS3 + .72*SWLS4 + .61*SWLS5

#Means
Authenticity ~ 4.30*1
Connection ~ 4.16*1
Home ~ 4.51*1
RECthriving ~ 4.81*1
SelfDefinition ~ 4.53*1
GBS ~ 4.84*1
CSESR ~ 5.15*1
MEIMR ~ 3.81*1
#CHS ~ 2.43*1
#ULS ~ 2.29*1
#MHI ~ 4.06*1
#SWLS ~ 4.09*1

#Correlations
Authenticity ~~ .73*Connection
Authenticity ~~ .78*Home
Authenticity ~~ .64*RECthriving
Authenticity ~~ .52*SelfDefinition
Authenticity ~~ .58*GBS
Authenticity ~~ .58*CSESR
Authenticity ~~ .48*MEIMR
#Authenticity ~~ -.44*CHS
#Authenticity ~~ -.43*ULS
```

```

#Authenticity ~~ .37*MHI
#Authenticity ~~ .34*SWLS

Connection ~~ .81*Home
Connection ~~ .70*RECthriving
Connection ~~ .43*SelfDefinition
Connection ~~ .36*GBS
Connection ~~ .63*CSESR
Connection ~~ .62*MEIMR
#Connection ~~ -.36*CHS
#Connection ~~ -.24*ULS
#Connection ~~ .18*MHI
#Connection ~~ .25*SWLS

Home ~~ .74*RECthriving
Home ~~ .48*SelfDefinition
Home ~~ .51*GBS
Home ~~ .65*CSESR
Home ~~ .63*MEIMR
#Home ~~ -.46*CHS
#Home ~~ -.34*ULS
#Home ~~ .27*MHI
#Home ~~ .29*SWLS

RECthriving ~~ .46*SelfDefinition
RECthriving ~~ .42*GBS
RECthriving ~~ .70*CSESR
RECthriving ~~ .70*MEIMR
#RECthriving ~~ -.32*CHS
#RECthriving ~~ -.31*ULS
#RECthriving ~~ .24*MHI
#RECthriving ~~ .30*SWLS

SelfDefinition ~~ .29*GBS
SelfDefinition ~~ .38*CSESR
SelfDefinition ~~ .35*MEIMR
#SelfDefinition ~~ -.30*CHS
#SelfDefinition ~~ -.28*ULS
#SelfDefinition ~~ .26*MHI
#SelfDefinition ~~ .22*SWLS
'

set.seed(240326)
iBel <- lavaan::simulateData(model = iBelong_generating_model,
                           model.type = "sem",
                           meanstructure = T,

```

```

sample.nobs=500,
standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(iBel))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 25 are the iBelong scale
#Rows 26 thru 37 are the General Belonging Scale
#Rows 38 thru 53 are the Collective Self Esteem Scale Race Specific Version
#Rows 54 thru 59 are the Multigroup Etnic Identity measure Revised

for(i in 1:ncol(iBel)){
  if(i >= 1 & i <= 25){
    iBel[,i] <- scales::rescale(iBel[,i], c(1, 6))
  }
  if(i >= 26 & i <= 37){
    iBel[,i] <- scales::rescale(iBel[,i], c(1, 7))
  }
  if(i >= 38 & i <= 53){
    iBel[,i] <- scales::rescale(iBel[,i], c(1, 7))
  }
  if(i >= 54 & i <= 59){
    iBel[,i] <- scales::rescale(iBel[,i], c(1, 5))
  }
}
#rounding to integers so that the data resembles that which was collected
library(tidyverse)
iBel <- iBel %>% round(0)

#quick check of my work
#psych::describe(iBel)

```

The optional script below will let you save the simulated data to your computing environment as either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(iBel, 'iBel.rds') bring back the simulated dat from an
# .rds file iBel <- readRDS('iBel.rds')

# write the simulated data as a .csv write.table(iBel,
# file='iBel.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file iBel <-
# read.csv('iBel.csv', header = TRUE)

```

Lessons in this volume teach test-retest reliability and invariance testing. In both of these circumstances, data from two different administrations of the focal test are required. For test-retest, these would be completed by the same person; for invariance-testing, these would be completed by different people. Below I have simulated two sets of data for the iBelong Scale [Lee and Neville, 2024]. I have named them “A” and “B”. For your homework purposes, you can determine if they will represent testing with the same individual (appropriate for test-retest reliability) or different groups (appropriate for invariance testing).

```
iBelong_generating_modelAB <- '
  #measurement model
  AuthenticityA =~ .66*Auth1a + .64*Auth2a + .63*Auth3a + .58*Auth4a + .55*Auth5a
  ConnectionA =~ .72*Conn1a + .60*Conn2a + .59*Conn3a + .58*Conn4a + .58*Conn5a
  HomeA =~ .64*Home1a + .63*Home2a + .57*Home3a + .57*Home4a + .53*Home5a
  RECthrivingA =~ .68*Thriv1a + .66*Thriv2a + .59*Thriv3a + .54*Thriv4a + .49*Thriv5a
  SelfDefinitionA =~ .82*Self1a + .65*Self2a + .65*Self3a + .54*Self4a + .49*Self5a
  AuthenticityB =~ .66*Auth1b + .64*Auth2b + .63*Auth3b + .58*Auth4b + .55*Auth5b
  ConnectionB =~ .72*Conn1b + .60*Conn2b + .59*Conn3b + .58*Conn4b + .58*Conn5b
  HomeB =~ .64*Home1b + .63*Home2b + .57*Home3b + .57*Home4b + .53*Home5b
  RECthrivingB =~ .68*Thriv1b + .66*Thriv2b + .59*Thriv3b + .54*Thriv4b + .49*Thriv5b
  SelfDefinitionB =~ .82*Self1b + .65*Self2b + .65*Self3b + .54*Self4b + .49*Self5b

  #Means
  AuthenticityA ~ 4.30*1
  ConnectionA ~ 4.16*1
  HomeA ~ 4.51*1
  RECthrivingA ~ 4.81*1
  SelfDefinitionA ~ 4.53*1

  AuthenticityB ~ 4.28*1
  ConnectionB ~ 4.20*1
  HomeB ~ 4.44*1
  RECthrivingB ~ 4.9*1
  SelfDefinitionB ~ 4.49*1

  #Correlations
  AuthenticityA ~~ .81*AuthenticityB
  ConnectionA ~~ .85*ConnectionB
  HomeA ~~ .77*HomeB
  RECthrivingA ~~ .85*RECthrivingB
  SelfDefinitionA ~~ .84*SelfDefinitionB

  '

set.seed(240326)
iBelAB <- lavaan::simulateData(model = iBelong_generating_modelAB,
                                model.type = "sem",
                                meanstructure = T,
```

```

        sample.nobs=500,
        standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(iBelAB))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#All rows are the iBel scales, administrations A and B

for(i in 1:ncol(iBelAB)){
  if(i >= 1 & i <= 50){
    iBelAB[,i] <- scales::rescale(iBelAB[,i], c(1, 6))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
iBelAB <- iBelAB %>% round(0)

#quick check of my work
#psych::describe(iBelAB)

```

## 11.2 Identity Threat

The Identity Threat Scale [George et al., 2023] is a measure of threat to identity value, meanings, and enactment. The measure is designed to adapt to different identities. The scale includes 19 items with responses rated on a 6 point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Items are written such that researchers can provide the focal identity (e.g., “a teacher,” “LGBTQ”). Across all three scales, higher scores reflect higher threat. It appears that the authors landed on a single-order, correlated factors, structure.

Beneath each of the three factors are a list of the items. Variable names follow in parentheses.

- Identity Value:
  - I feel that there is a negative value attached to my identity as [LGBTQ]. (Val1)
  - Being [LGBTQ] is worth less in the eyes of others than before. (Val2)
  - I feel that others attach a negative value to my identity as [LGBTQ]. (Val3)
  - I feel that my identity as [LGBTQ] is devalued by others. (Val4)
  - I feel that others see little value in my identity as [LGBTQ].(Val5)
- Identity Meaning:
  - I am no longer sure what it means to be [LGBTQ].(Mng1)
  - I am questioning what it means to be [LGBTQ]. (Mng2)
  - I find myself questioning what it means to be [LGBTQ]. (Mng3)
  - The core of what it means to be [LGBTQ] is changing in a way I do not like. (Mng4)

- What it means to be [LGBTQ] is changing in a way I do not like. (Mng5)
  - What it means to be [LGBTQ] is being called into question. (Mng6)
  - Being [LGBTQ] used to mean something different. (Mng7) -I feel that being [LGBTQ] does not mean the same thing anymore. (Mng8)
- Identity Enactment
    - I am limited in my ability to express my identity as [LGBTQ]. (Enact1)
    - I may no longer be able to engage in activities that express my identity as [LGBTQ]. (Enact2)
    - I may no longer be able to show that I am [LGBTQ]. (Enact3)
    - I worry about no longer being able to express my identity as [LGBTQ]. (Enact4)
    - I worry that I cannot behave in the way [an LGBTQ person] behaves. (Enact5)
    - I worry that I cannot show people that I am [LGBTQ]. (Enact6)

Although I have simulated data, the authors posted their data and codebooks from the different phases at the [OSF repository](#). Additional materials are available as a [supplement](#) to the article.

There were a number of scales utilized in this study. Because including them all caused problems of convergence, beyond the subscales in the focal measure, I have simulated data for only a few more. This will provide practice in evaluating convergent, discriminant, and incremental validity.

- **Self-esteem** was assessed with [Rosenberg's 10-item scale](#). Responses are measured on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). Some items are reverse-coded. However, using the data from this simulation presumes that reverse-scoring has already been completed. Higher scores indicate higher self-esteem. An example item is “On the whole, I am satisfied with myself.” Factor loadings were obtained from Mullen et al. [\[2013\]](#).
- **Identity suppression** was assessed with 4 items from Madera et al.’s [\[2012\]](#) 10 item scale. It is a bit unclear, but it appears that items were rated on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale with higher scores reflecting higher identity suppression. A sample item is, “I refrain from talking about my identity as LGBTQ at work.”
- **Workplace microaggressions** was captured with 11 items from Resnick and Galupo’s [\[Resnick and Galupo, 2019\]](#) scale. Responses are measured on a Likert type scale ranging from 1 (*never*) to 5 (*a great deal*). Higher scores indicate higher frequency of microaggressions. An example item is “Having my behaviors mimicked in a joking way due to my LGBTQ identity.” It is unclear if they used one or more scales. I simulated the data from Resnick and Galupo, selecting 3 and 4 items from three subscales.
- **Green behaviors** was captured with three items from Norton et al.’s [\[Norton et al., 2017\]](#) scale. Responses are measured on a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The response stem was, “Tomorrow, I intend to...” and items included, “...act in environmentally friendly ways,” “...carry out environmentally friendly behaviors at work,” and “...perform pro-environmental behaviors while at work. In the absence of factor loadings, I made them up.

Because data is collected at the item level (and I want this resource to be as practical as possible, I have simulated the data for each of the scales at the item level. I guided the simulation using factor loadings from the exploratory factor analysis results and means and correlations from the Stage 5 analysis where 516 individuals who identified as LGBTQ completed the survey. Because

the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

*#Entering the intercorrelations, means, and standard deviations from the journal article*

```
IdentityThreat_generating_model <- '
  #measurement model
  Value  =~ .83*Val1 + .78*Val2 + .88*Val3 + .80*Val4 + .73*Val5
  Meaning =~ .66*Mng1 + .81*Mng2 + .85*Mng3 + .78*Mng4 + .74*Mng5 + .64*Mng6 + .65*Mng7 +
  Enactment =~ .66*Enact1 + .64*Enact2 + .82*Enact3 + .70*Enact4 + .66*Enact5 + .94*Enact6
  SelfEsteem =~ .95*SE1 + .98*SE2 + .66*SE3 + .81*SE4 + .55*SE5 + .80*SE6 + .75*SE7 + .68*SE8
  IDs suppression =~ .78*IDs1 + .67*IDs2 + .82*IDs3 + .90*IDs4
  WkplcMicroAgg =~ .84*Micro1 + .87*Micro2 + .72*Micro3 + .78*Micro4 + .67*Micro5 + .81*Micro6
  GreenBx =~ .77*G1 + .81*G2 + .83*G3

  #Means
  Value ~ 2.15*1
  Meaning ~ 2.10*1
  Enactment ~ 2.17*1
  SelfEsteem ~ 3.44*1
  IDs suppression ~ 3.13*1
  WkplcMicroAgg ~ 1.24*1
  GreenBx ~ 3.49*1

  #Correlations
  Value ~~ .58*Meaning
  Value ~~ .73*Enactment
  Value ~~ -.21*SelfEsteem
  Value ~~ .35*IDs suppression
  Value ~~ .41*WkplcMicroAgg
  Value ~~ -.09*GreenBx

  Meaning ~~ .60*Enactment
  Meaning ~~ .63*SelfEsteem
  Meaning ~~ .26*IDs suppression
  Meaning ~~ .24*WkplcMicroAgg
  Meaning ~~ -.08*GreenBx

  Enactment ~~ -.31*SelfEsteem
  Enactment ~~ .54*IDs suppression
  Enactment ~~ .24*WkplcMicroAgg
  Enactment ~~ -.06*GreenBx

  SelfEsteem ~~ -.24*IDs suppression
  SelfEsteem ~~ -.06*WkplcMicroAgg
  SelfEsteem ~~ .15*GreenBx

  IDs suppression ~~ .00*WkplcMicroAgg
```

```

IDsuppression ~~ -.02*GreenBx

set.seed(240330)
IdThreat <- lavaan::simulateData(model = IdentityThreat_generating_model,
                                   model.type = "sem",
                                   meanstructure = T,
                                   sample.nobs=516,
                                   standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(IdThreat))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 19 are the Identity Threat
#Rows 20 thru 29 are Self Esteem
#Rows 30 thru 33 are Identity Suppression
#Rows 34 thru 44 are Workplace Microaggressions
#Rows 45 thru 47 are Green Behaviors

for(i in 1:ncol(IdThreat)){
  if(i >= 1 & i <= 19){
    IdThreat[,i] <- scales::rescale(IdThreat[,i], c(1, 7))
  }
  if(i >= 20 & i <= 29){
    IdThreat[,i] <- scales::rescale(IdThreat[,i], c(1, 4))
  }
  if(i >= 30 & i <= 33){
    IdThreat[,i] <- scales::rescale(IdThreat[,i], c(1, 7))
  }
  if(i >= 34 & i <= 44){
    IdThreat[,i] <- scales::rescale(IdThreat[,i], c(1, 5))
  }
  if(i >= 45 & i <= 47){
    IdThreat[,i] <- scales::rescale(IdThreat[,i], c(1, 5))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
IdThreat <- IdThreat %>% round(0)

#quick check of my work
#psych::describe(IdThreat)

```

The optional script below will let you save the simulated data to your computing environment as

either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```
# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(IdThreat, 'IdThreat.rds') bring back the simulated dat
# from an .rds file IdThreat <- readRDS('IdThreat.rds')
```

```
# write the simulated data as a .csv write.table(IdThreat,
# file='IdThreat.csv', sep=',', col.names=TRUE, row.names=FALSE)
# bring back the simulated dat from a .csv file IdThreat <-
# read.csv('IdThreat.csv', header = TRUE)
```

Lessons in this volume teach test-retest reliability and invariance testing. In both of these circumstances, data from two different administrations of the focal test are required. For test-retest, these would be completed by the same person; for invariance-testing, these would be completed by different people. Below I have simulated two sets of data for The Identity Threat Scale [George et al., 2023]. I have named them “A” and “B”. For your homework purposes, you can determine if they will represent testing with the same individual (appropriate for test-retest reliability) or different groups (appropriate for invariance testing).

```
idThreat_generating_modelAB <- '
  #measurement model
  ValueA =~ .83*Val1a + .78*Val2a + .88*Val3a + .80*Val4a + .73*Val5a
  MeaningA =~ .66*Mng1a + .81*Mng2a + .85*Mng3a + .78*Mng4a + .74*Mng5a + .64*Mng6a + .68*Mng7a
  EnactmentA =~ .66*Enact1a + .64*Enact2a + .82*Enact3a + .70*Enact4a + .66*Enact5a + .94*Enact6a
  ValueB =~ .83*Val1b + .78*Val2b + .88*Val3b + .80*Val4b + .73*Val5b
  MeaningB =~ .66*Mng1b + .81*Mng2b + .85*Mng3b + .78*Mng4b + .74*Mng5b + .64*Mng6b + .68*Mng7b
  EnactmentB =~ .66*Enact1b + .64*Enact2b + .82*Enact3b + .70*Enact4b + .66*Enact5b + .94*Enact6b

  #Means
  ValueA ~ 1.97*1
  MeaningA ~ 2.16*1
  EnactmentA ~ 2.14*1
  ValueB ~ 2.10*1
  MeaningB ~ 2.28*1
  EnactmentB ~ 2.24*1

  #Correlations
  ValueA ~~ .80*MeaningA
  ValueA ~~ .80*EnactmentA
  ValueA ~~ .63*ValueB
  ValueA ~~ .57*MeaningB
  ValueA ~~ .56*EnactmentB

  MeaningA ~~ .77*EnactmentA
  MeaningA ~~ .59*ValueB
```

```

MeaningA ~~ .70*MeaningB
MeaningA ~~ .58*EnactmentB

EnactmentA ~~ .54*ValueB
EnactmentA ~~ .79*MeaningB
EnactmentA ~~ .66*EnactmentB

ValueB ~~ .81*MeaningB
ValueB ~~ .82*EnactmentB

MeaningB ~~ .81*EnactmentB

'

set.seed(240326)
idThreatAB <- lavaan::simulateData(model = idThreat_generating_modelAB,
                                      model.type = "sem",
                                      meanstructure = T,
                                      sample.nobs=500,
                                      standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(idThreatAB))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#All rows are the iBel scales, administrations A and B

for(i in 1:ncol(idThreatAB)){
  if(i >= 1 & i <= 38){
    idThreatAB[,i] <- scales::rescale(idThreatAB[,i], c(1, 5))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
idThreatAB <- idThreatAB %>% round(0)

#quick check of my work
#psych::describe(iBelAB)

```

## 11.3 Anti-Racism Behavioral Inventory

The Anti-Racism Behavioral Inventory [Pieterse et al., 2022] is a measure that assesses anti-racism awareness and behavior among white Americans. The scale includes 21 items with responses rated on a 5 point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Higher scores

indicate a greater anti-racist awareness and behavior.

Beneath each of the five factors are a list of the items. Variable names follow in parentheses.

- Individual Advocacy:

- When I hear people telling racist jokes and using negative racial stereotypes, I usually confront them (Advoc1)
- I actively seek to understand how I participate in both intentional and unintentional racism (Advoc2)
- I actively seek to educate myself about the experience of racism (Advoc3)
- I interrupt racist conversations and jokes when I hear my friends talking that way (Advoc4)
- I have challenged acts of racism that I have witnessed in my workplace or at school (Advoc5)
- I make it a point to educate myself about the experience of historically oppressed groups in the US (e.g. slavery, internment of Japanese, American Indians and the trail of tears etc.) (Advoc6)
- I often speak to my friends about the problem of racism in the US, and what we can do about it (Advoc7)
- I do not like to talk about racism in public (Advoc8) NEEDS REVERSING
- I interrupt racist conversations and jokes when I hear them in my family (Advoc9)

- Awareness of Racism:

- I feel guilty and ashamed when I think of the history of racism and slavery in the US (Aware1)
- It bothers me that my country has yet to acknowledge the impact of slavery (Aware2)
- The US should offer some type of payment to the descendants of slaves (Aware3)
- The US has not acknowledged the impact of slavery (Aware4)
- Because of racism in the US, Blacks do not have the same educational opportunities as compared to Whites (Aware5)
- Within the US, racism is largely perpetuated by the White racial majority (Aware6)
- The police unfairly target Black men and Latinos (Aware7)

- Institutional Advocacy

- I give money to organizations working against racism and discrimination (Inst1)
- When I read articles in newspapers or magazines that are perpetuating racist ideas, I generally write a letter to the editor (Inst2)
- I am actively involved in exposing companies that uphold exclusionary and racist practices (Inst3)
- I write letters to local and state politicians to voice my concerns about racism (Inst4)
- I volunteer with anti-racist or racial justice organizations (Inst5)

This article was published in the open access Journal for Social Action in Counseling and Psychology and is available at its [website](#). Below is the script for simulating item-level data from the factor loadings, means, and sample size presented in the published article.

Four additional scales were reported in the Pieterse et al. [2022] article. Unfortunately, I could not locate factor loadings for all of them; in these cases I uniformly assigned the factor loading of 0.8.

The scales, their original citation, and information about how I simulated data for each are listed below.

- **White Privilege** was assessed using a 13-item subscale from The Privilege and Oppression Inventory\*\* [Hays et al., 2007]. Items are rated on a Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). An example item is, “Whites have the power to exclude other groups.” Higher scores indicate greater levels of White privilege.
- **Color Blind Racial Attitudes** [Neville et al., 2000] is a 20-item scale that assesses implicit racism-related attitudes. The 21 items are assessed on a Likert scale ranging from 1 to 5 with higher scores reflecting greater levels of color-blind racial attitudes. I was able to simulate two of the three subscales of the instrument: Racial Privilege (RP) and Blatant Racial Issues (BR).
- **Mental Health Inventory-5** [Berwick et al., 1991] is a five-item mental health screening tool. Responses are made on a Likert scale ranging from 1 (*all of the time*) to 5 (*none of the time*). Higher scores reflect greater mental health.

Because data is collected at the item level (and I want this resource to be as practical as possible, I have simulated the data for each of the scales at the item level. Simulating the data involved using factor loadings, means, and correlations between the scales. Where possible, factor loadings are retrieved from original (or replication) articles that report on their psychometric evaluation. The means and correlations are from Pietierse et al. [Pieterse et al., 2022] Because the simulation will produce “out-of-bounds” values, the code below rescales the scores into the range of the Likert-type scaling and rounds them to whole values.

*#Entering the intercorrelations, means, and standard deviations from the journal article*

```
Antiracism_generating_model <- '
  #measurement model
  ARBI =~ .274*Advoc1 + .699*Advoc2 + .752*Advoc3 + .465*Advoc4 + .479*Advoc5 + .617*Advoc6
  Advocacy =~ .618*Advoc1 + .055*Advoc2 + .035*Advoc3 + .712*Advoc4 + .275*Advoc5 + .124*Advoc6
  Awareness =~ .211*Aware1 + .669*Aware2 + .428*Aware3 + .692*Aware4 + .445*Aware5 + .509*Aware6
  Institutional =~ .381*Inst1 + .547*Inst2 + .648*Inst3 + .618*Inst4 + .392*Inst5
  WhitePriv =~ .80*WP1 + .59*WP2 + .62*WP3 + .65*WP4 + .80*WP5 + .50*WP6 + .78*WP7 + .57*WP8
  CoBRAS_RP =~ .68*RP1 + .66*RP2 + .64*RP3 + .62*RP4 + .62*RP5 + .58*RP6 + .54*RP7
  CoBRAS_BR =~ .77*BR1 + .67*BR2 + .65*BR3 + .65*BR4 + .62*BR5 + .55*BR6 + .48*BR7
  #CoBRAS_ID =~ .65*ID1 + .64*ID2 + .61*ID3 + .56*ID4 + .56*ID5 + .56*ID6 + .53*ID7

  #Means
  ARBI ~ 3.91*1
  Advocacy ~5.33*1
  Awareness ~ 4.26*1
  Institutional ~ 1.66*1
  WhitePriv ~ 5.20*1
  CoBRAS_RP ~ 2.54*1
  CoBRAS_BR ~ 1.65*1
  #CoBRAS_ID ~ 2.00*1
```

```

#Correlations
ARBI ~~ .78*Awareness
ARBI ~~ .87*Advocacy
ARBI ~~ .75*Institutional
ARBI ~~ .64*WhitePriv
ARBI ~~ -.65*CoBRAS_RP
#ARBI ~~ -.70*CoBRAS_ID
ARBI ~~ -.60*CoBRAS_BR

Awareness ~~ .50*Advocacy
Awareness ~~ .34*Institutional
Awareness ~~ .81*WhitePriv
Awareness ~~ -.78*CoBRAS_RP
#Awareness ~~ -.76*CoBRAS_ID
Awareness ~~ -.62*CoBRAS_BR

Advocacy ~~ .55*Institutional
Advocacy ~~ .43*WhitePriv
Advocacy ~~ -.45*CoBRAS_RP
#Advocacy ~~ -.50*CoBRAS_ID
Advocacy ~~ -.46*CoBRAS_BR

Institutional ~~ .27*WhitePriv
Institutional ~~ -.31*CoBRAS_RP
#Institutional ~~ -.39*CoBRAS_ID
Institutional ~~ -.34*CoBRAS_BR

WhitePriv ~~ -.85*CoBRAS_RP
#WhitePriv ~~ -.8*CoBRAS_ID
WhitePriv ~~ -.6*CoBRAS_BR

'

set.seed(240218)
AntiR <- lavaan:::simulateData(model = Antiracism_generating_model,
                                model.type = "sem",
                                meanstructure = T,
                                sample.nobs=153,
                                standardized=FALSE)

#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(AntiR))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#Rows 1 thru 21 are the ARBI
#Rows 22 thru 34 are the White Privilege Scale
#Rows 35 thru 48 are the CoBRAS

```

```

for(i in 1:ncol(AntiR)){
  if(i >= 1 & i <= 21){
    AntiR[,i] <- scales::rescale(AntiR[,i], c(1, 5))
  }
  if(i >= 22 & i <= 34){
    AntiR[,i] <- scales::rescale(AntiR[,i], c(1, 6))
  }
  if(i >= 35 & i <= 48){
    AntiR[,i] <- scales::rescale(AntiR[,i], c(1, 5))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
AntiR <- AntiR %>% round(0)

#quick check of my work
#psych::describe(AntiR)

#Reversing the supportive item on the Perceptions of LGBTQ Campus Climate Scale so that the ex

```

The optional script below will let you save the simulated data to your computing environment as either an .rds object (preserves any formatting you might do) or a.csv file (think “Excel lite”).

```

# to save the df as an .rds (think 'R object') file on your computer;
# it should save in the same file as the .rmd file you are working
# with saveRDS(AntiR, 'AntiR.rds') bring back the simulated dat from
# an .rds file AntiR <- readRDS('AntiR.rds')

# write the simulated data as a .csv write.table(AntiR,
# file='AntiR.csv', sep=',', col.names=TRUE, row.names=FALSE) bring
# back the simulated dat from a .csv file
AntiR <- read.csv("AntiR.csv", header = TRUE)

```

Lessons in this volume teach test-retest reliability and invariance testing. In both of these circumstances, data from two different administrations of the focal test are required. For test-retest, these would be completed by the same person; for invariance-testing, these would be completed by different people. Below I have simulated two sets of data for the Anti-Racism Behavioral Inventory [Pieterse et al., 2022]. I have named them “A” and “B”. For your homework purposes, you can determine if they will represent testing with the same individual (appropriate for test-retest reliability) or different groups (appropriate for invariance testing).

The Pieterse et al.[2022] article did not report invariance testing nor test-retest reliabilities. Therefore, I simulated this data following the general patterns of such testing.

```

ARBI_generating_modelAB <- '
  #measurement model
  AdvocacyA =~ .450*Advoc1a + .671*Advoc2a + .715*Advoc3a + .650*Advoc4a + .548*Advoc5a
  AwarenessA =~ .239*Aware1a + .836*Aware2a + .754*Aware3a + .877*Aware4a + .791*Aware5a
  InstitutionalA =~ .698*Inst1a + .397*Inst2a + .680*Inst3a + .589*Inst4a + .798*Inst5a
  AdvocacyB =~ .450*Advoc1b + .671*Advoc2b + .715*Advoc3b + .650*Advoc4b + .548*Advoc5b
  AwarenessB =~ .239*Aware1b + .836*Aware2b + .754*Aware3b + .877*Aware4b + .791*Aware5b
  InstitutionalB =~ .698*Inst1b + .397*Inst2b + .680*Inst3b + .589*Inst4b + .798*Inst5b

  #Means
  AdvocacyA ~5.33*1
  AwarenessA ~ 4.26*1
  InstitutionalA ~ 1.66*1
  AdvocacyB ~5.25*1
  AwarenessB ~ 4.30*1
  InstitutionalB ~ 1.71*1

  #Correlations
  AdvocacyA ~~ .80*AwarenessA
  AdvocacyA ~~ .80*InstitutionalA
  AdvocacyA ~~ .63*AdvocacyB
  AdvocacyA ~~ .57*AwarenessB
  AdvocacyA ~~ .56*InstitutionalB

  AwarenessA ~~ .77*InstitutionalA
  AwarenessA ~~ .59*AdvocacyB
  AwarenessA ~~ .70*AwarenessB
  AwarenessA ~~ .58*InstitutionalB

  InstitutionalA ~~ .54*AdvocacyB
  InstitutionalA ~~ .79*AwarenessB
  InstitutionalA ~~ .66*InstitutionalB

  AdvocacyB ~~ .81*AwarenessB
  AdvocacyB ~~ .82*InstitutionalB

  AwarenessB ~~ .81*InstitutionalB

  '

set.seed(240326)
ARBI_AB <- lavaan::simulateData(model = ARBI_generating_modelAB,
                                 model.type = "sem",
                                 meanstructure = T,
                                 sample.nobs=300,
                                 standardized=FALSE)

```

```
#used to retrieve column indices used in the rescaling script below
col_index <- as.data.frame(colnames(ARBI_AB))

#The code below loops through each column of the dataframe and assigns the scaling accordingly
#All rows are the iBel scales, administrations A and B

for(i in 1:ncol(ARBI_AB)){
  if(i >= 1 & i <= 38){
    ARBI_AB[,i] <- scales::rescale(ARBI_AB[,i], c(1, 5))
  }
}

#rounding to integers so that the data resembles that which was collected
library(tidyverse)
ARBI_AB <- ARBI_AB %>% round(0)

#quick check of my work
#psych::describe(iBelAB)
```



# REFERENCES



# Bibliography

FACT SHEET: Anti-Asian Prejudice March 2020, a. URL <https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-Asian%20Hate%202020%203.2.21.pdf>.

STOP AAPI HATE, b. URL <https://stopaapihate.org/>.

Hector Y. Adames, Nayeli Y. Chavez-Dueñas, and Maryam M. Jernigan. The fallacy of a raceless Latinidad: Action guidelines for centering Blackness in Latinx psychology. *Journal of Latinx Psychology*, 9(1):26–44, February 2021. ISSN 2578-8086. doi: 10.1037/lat0000179. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2020-69785-001&site=ehost-live>. Publisher: Educational Publishing Foundation.

Anne Anastasi and Susana Urbina. *Psychological testing*, 7th ed. Prentice Hall/Pearson Education, Upper Saddle River, NJ, 1997. ISBN 978-0-02-303085-7. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=1998-07223-000&site=ehost-live>.

A Bandura and Albert Bandura. Guide for creating self-efficacy scales. In F. Pajares and T. Urdan, editors, *Self-efficacy beliefs of adolescents*, pages 307–337. Information Age Publishing, Greenwich, CT, 2006.

D. M. Berwick, J. M. Murphy, P. A. Goldman, J. E. Ware, A. J. Barsky, and M. C. Weinstein. Performance of a five-item mental health screening test. *Medical Care*, 29(2):169–176, February 1991. ISSN 0025-7079. doi: 10.1097/00005650-199102000-00008.

Martin Brattmyr, Martin Schevik Lindberg, Stian Solem, Odin Hjemdal, and Audun Havnen. Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder Scale-7 in a Norwegian psychiatric outpatient sample. *BMC Psychiatry*, 22(1):461, July 2022. ISSN 1471-244X. doi: 10.1186/s12888-022-04101-z. URL <https://doi.org/10.1186/s12888-022-04101-z>.

Susan D. Brown, Kirsten A. Unger Hu, Ashley A. Mevi, Monique M. Hedderson, Jun Shan, Charles P. Quesenberry, and Assiamira Ferrara. The Multigroup Ethnic Identity Measure-Revised: Measurement invariance across racial and ethnic groups. *Journal of counseling psychology*, 61(1):154–161, January 2014. ISSN 0022-0167. doi: 10.1037/a0034749. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110058/>.

Elizabeth A. Buchanan and Erin E. Hvizdak. Online Survey Tools: Ethical and Methodological Concerns of Human Research Ethics Committees. *Journal of Empirical Research on Human Research Ethics; London*, 4(2):37–48, June 2009. ISSN 15562646. doi: <http://dx.doi.org/>

- ezproxy.spu.edu/10.1525/jer.2009.4.2.37. URL <http://search.proquest.com/docview/219055147/abstract/BE05412DDB224725PQ/9>.
- Barbara M. Byrne. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, Third Edition*. Taylor & Francis Group, London, UNITED KINGDOM, 2016. ISBN 978-1-317-63313-6. URL <http://ebookcentral.proquest.com/lib/spu/detail.action?docID=4556523>.
- Cristalis Capielo Rosario, Hector Y. Adames, Nayeli Y. Chavez-Dueñas, and Roberto Renteria. Acculturation Profiles of Central Florida Puerto Ricans: Examining the Influence of Skin Color, Perceived Ethnic-Racial Discrimination, and Neighborhood Ethnic-Racial Composition. *Journal of Cross-Cultural Psychology*, 50(4):556–576, May 2019. ISSN 0022-0221, 1552-5422. doi: 10.1177/0022022119835979. URL <http://journals.sagepub.com/doi/10.1177/0022022119835979>.
- Andrew Young Choi, Tania Israel, and Hotaka Maeda. Development and evaluation of the Internalized Racism in Asian Americans Scale (IRAAS): Journal of Counseling Psychology. *Journal of Counseling Psychology*, 64(1):52–64, January 2017. ISSN 0022-0167. doi: 10.1037/cou0000183. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2017-00131-004&site=ehost-live>. Publisher: American Psychological Association.
- Seung Youn Chyung, Katherine Roberts, Ieva Swanson, and Andrea Hankinson. Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, 56(10):15–23, November 2017. ISSN 10908811. doi: 10.1002/pfi.21727. URL <http://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=eue&AN=126315473&site=ehost-live>.
- Seung Youn Chyung, Julie R. Barkin, and Jennifer A. Shamsy. Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys. *Performance Improvement*, 57(3):16–25, March 2018a. ISSN 10908811. doi: 10.1002/pfi.21749. URL <http://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=eue&AN=128710204&site=ehost-live>.
- Seung Youn Chyung, Megan Kennedy, and Ingrid Campbell. Evidence-Based Survey Design: The Use of Ascending or Descending Order of Likert-Type Response Options. *Performance Improvement*, 57(9):9–16, 2018b. ISSN 1930-8272. doi: 10.1002/pfi.21800. URL <http://onlinelibrary.wiley.com/doi/abs/10.1002/pfi.21800>.
- Seung Youn Chyung, Ieva Swanson, Katherine Roberts, and Andrea Hankinson. Evidence-Based Survey Design: The Use of Continuous Rating Scales in Surveys. *Performance Improvement*, 57(5):38–48, May 2018c. ISSN 10908811. doi: 10.1002/pfi.21763. URL <http://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=bth&AN=129411016&site=ehost-live>.
- David Colton and Robert W. Covert. *Designing and Constructing Instruments for Social Research and Evaluation*. John Wiley & Sons, June 2015. ISBN 978-1-119-17797-5. Google-Books-ID: MLD5CQAAQBAJ.
- Jose M. Cortina. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1):98–104, 1993. ISSN 0021-9010. doi: 10.1037/0021-9010.78.1.98. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-9010.78.1.98>.

- Jennifer Crocker, Riia Luhtanen, Bruce Blaine, and Stephanie Broadnax. Collective self-esteem and psychological well-being among White, Black, and Asian college students: Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 20(5):503–513, October 1994. ISSN 0146-1672. doi: 10.1177/0146167294205007. Publisher: Sage Publications.
- Ed Diener, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. The Satisfaction With Life Scale: Journal of Personality Assessment. *Journal of Personality Assessment*, 49(1):71–75, February 1985. ISSN 0022-3891. doi: 10.1207/s15327752jpa4901\_13. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=1985-27000-001&site=ehost-live>. Publisher: Lawrence Erlbaum.
- Andy P. Field. *Discovering statistics using R*. Sage, Thousand Oaks, California, 2012. ISBN 978-1-4462-0046-9.
- Mailys M. George, Karoline Strauss, Julija N. Mell, and Heather C. Vough. When 'who I am' is under threat: Measures of threat to identity value, meanings, and enactment: Journal of Applied Psychology. *Journal of Applied Psychology*, 108(12):1952–1978, December 2023. ISSN 0021-9010. doi: 10.1037/apl0001114. supp. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2023-93532-001&site=ehost-live>. Publisher: American Psychological Association.
- Malcolm Gladwell. *Outliers: the story of success*. New York Times best sellers. Little, Brown and Company, New York, first edition. edition, 2008. ISBN 978-0-316-01792-3.
- Samuel B. Green and Neil J. Salkind. *Using SPSS for Windows and Macintosh: analyzing and understanding data*. Pearson, Boston, eighth edition. edition, 2017. ISBN 978-0-13-431988-9.
- Danica G. Hays, Catherine Y. Chang, and Scott L. Decker. Initial development and psychometric data for the Privilege and Oppression Inventory: Measurement and Evaluation in Counseling and Development. *Measurement and Evaluation in Counseling and Development*, 40(2):66–79, July 2007. ISSN 0748-1756. doi: 10.1080/07481756.2007.11909806. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2007-10909-001&site=ehost-live>. Publisher: American Counseling Assn.
- Ron D. Hays and M. Robin DiMatteo. A short-form measure of loneliness: Journal of Personality Assessment. *Journal of Personality Assessment*, 51(1):69–81, 1987. ISSN 0022-3891. doi: 10.1207/s15327752jpa5101\_6. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=1987-20763-001&site=ehost-live>. Publisher: Lawrence Erlbaum.
- Edward G. Helm, William E. Sedlacek, and Dario O. Prieto. The Relationship Between Attitudes Toward Diversity and Overall Satisfaction of University Students by Race. *Journal of College Counseling*, 1(2):111–120, September 1998. ISSN 1099-0399, 2161-1882. doi: 10.1002/j.2161-1882.1998.tb00130.x. URL <https://onlinelibrary.wiley.com/doi/10.1002/j.2161-1882.1998.tb00130.x>.
- G. M. Herek. Documenting prejudice against lesbians and gay men on campus: the Yale Sexual Orientation Survey. *Journal of Homosexuality*, 25(4):15–30, 1993. ISSN 0091-8369. doi: 10.1300/J082v25n04\_02.

- Clara E. Hill. *Helping skills: facilitating exploration, insight, and action*. American Psychological Association, Washington, DC, fifth edition. edition, 2020. ISBN 978-1-4338-3137-9.
- Sylvia Hurtado. Linking Diversity with the Educational and Civic Missions of Higher Education. *Review of Higher Education: Journal of the Association for the Study of Higher Education*, 30(2):185–196, 2007. ISSN 0162-5748. doi: 10.1353/rhe.2006.0070. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2006-23268-004&site=ehost-live>. Publisher: Johns Hopkins University Press.
- Juliette M. Iacobino and Sherman A. James. Retaining Students of Color in Higher Education: Expanding Our Focus to Psychosocial Adjustment and Mental Health. In *The Crisis of Race in Higher Education: A Day of Discovery and Dialogue*, volume 19 of *Diversity in Higher Education*, pages 61–84. Emerald Group Publishing Limited, January 2016. ISBN 978-1-78635-710-6 978-1-78635-709-0. doi: 10.1108/S1479-364420160000019004. URL <https://doi.org/10.1108/S1479-364420160000019004>.
- Rajiv S. Jhangiani, I.-Chant A. Chiang, Carrie Cuttler, and Dana C. Leighton. *Research Methods in Psychology*. August 2019. ISBN 978-1-9991981-0-7. doi: 10.17605/OSF.IO/HF7DQ. URL <https://kpu.pressbooks.pub/psychmethods4e/>.
- Brian TaeHyuk Keum, Jennifer L. Brady, Rajni Sharma, Yun Lu, Young Hwa Kim, and Christina J. Thai. Gendered Racial Microaggressions Scale for Asian American Women: Development and initial validation. *Journal of Counseling Psychology*, 65(5):571–585, October 2018. ISSN 0022-0167. doi: 10.1037/cou0000305. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2018-36321-001&site=ehost-live>.
- Paul Kim. Yes, Asians and Asian Americans experience racism, March 2021a. URL <https://www.seattletimes.com/opinion/yes-asians-and-asian-americans-experience-racism/>. Section: Opinion.
- Paul Y. Kim. Guest Post: Anti-Asian Racism during the Pandemic: How Faculty on Christian Campuses Can Support Asian and Asian American Students, March 2021b. URL <https://christianscholars.com/guest-post-anti-asian-racism-during-the-pandemic-how-faculty-on-christian-campuses-can-support-asian-and-asian-american-students/>.
- Rex B. Kline. *Principles and practice of structural equation modeling*. Guilford Publications, New York, UNITED STATES, 4th edition, 2016. ISBN 978-1-4625-2336-8. URL <http://ebookcentral.proquest.com/lib/spu/detail.action?docID=4000663>.
- Elizabeth A. Klonoff and Hope Landrine. The Schedule of Sexist Events: A measure of lifetime and recent sexist discrimination in women's lives. *Psychology of Women Quarterly*, 19(4):439–472, December 1995. ISSN 0361-6843. doi: 10.1111/j.1471-6402.1995.tb00086.x. Publisher: Blackwell Publishing.
- David R. Krathwohl. *Methods of Educational and Social Science Research: The Logic of Methods, Third Edition*. Waveland Press, February 2009. ISBN 978-1-4786-3425-6. Google-Books-ID: 0lMbDQAQBAJ.
- K. Kroenke, R. L. Spitzer, and J. B. Williams. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606–613, September 2001. ISSN 0884-8734. doi: 10.1046/j.1525-1497.2001.016009606.x.

- B. Andi Lee and Helen A. Neville. The iBelong Scale: Construction and validation of a measure of racial–ethnic–cultural belonging. *Journal of Counseling Psychology*, 71(3):139–154, April 2024. ISSN 0022-0167. doi: 10.1037/cou0000716. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2024-48470-001&site=ehost-live>. Publisher: American Psychological Association.
- Jioni A. Lewis and Helen A. Neville. Construction and initial validation of the Gendered Racial Microaggressions Scale for Black women. *Journal of Counseling Psychology*, 62(2):289–302, April 2015. ISSN 0022-0167. doi: 10.1037/cou0000062. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2015-15500-004&site=ehost-live>. Publisher: American Psychological Association.
- Kaleea R. Lewis and Payal P. Shah. Black students' narratives of diversity and inclusion initiatives and the campus racial climate: An interest-convergence analysis. *Journal of Diversity in Higher Education*, October 2019. ISSN 1938-8926. doi: 10.1037/dhe0000147. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2019-59378-001&site=ehost-live>. Publisher: Educational Publishing Foundation.
- Juan M. Madera, Eden B. King, and Michelle R. Hebl. Bringing social identity to work: The influence of manifestation and suppression on perceived discrimination, job satisfaction, and turnover intentions. *Cultural Diversity and Ethnic Minority Psychology*, 18(2):165–170, April 2012. ISSN 1099-9809. doi: 10.1037/a0027724. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2012-09819-006&site=ehost-live>. Publisher: Educational Publishing Foundation.
- Brent Mallinckrodt, Joseph R. Miles, and Jacob J. Levy. The scientist-practitioner-advocate model: Addressing contemporary training needs for social justice advocacy. *Training and Education in Professional Psychology*, 8(4):303–311, November 2014. ISSN 1931-3918. doi: 10.1037/tep0000045. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2014-25072-001&site=ehost-live>. Publisher: Educational Publishing Foundation.
- Glenn P. Malone, David R. Pillow, and Augustine Osman. The General Belongingness Scale (GBS): Assessing achieved belongingness. *Personality and Individual Differences*, 52(3):311–316, February 2012. ISSN 0191-8869. doi: 10.1016/j.paid.2011.10.027. URL <https://www.sciencedirect.com/science/article/pii/S019188691100482X>.
- Daniel McNeish. Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3):412–433, September 2018. ISSN 1082-989X. doi: 10.1037/met0000144. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2017-23572-001&site=ehost-live>. Publisher: American Psychological Association.
- Samuel Messick. Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. In Richard D. Goffin and Edward Helmes, editors, *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy.*, pages 3–20. Kluwer

- Academic/Plenum Publishers, New York, NY, 2000. ISBN 978-0-7923-7768-9. doi: 10.1007/978-1-4615-4397-8\_1. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2001-16052-001&site=ehost-live>.
- Della V. Mosley, Helen A. Neville, Nayeli Y. Chavez-Dueñas, Hector Y. Adames, Jioni A. Lewis, and Bryana H. French. Radical hope in revolting times: Proposing a culturally relevant psychological framework. *Social and Personality Psychology Compass*, 14(1), January 2020. ISSN 1751-9004. doi: 10.1111/spc3.12512. Publisher: Wiley-Blackwell Publishing Ltd.
- Della V. Mosley, Candice N. Hargons, Carolyn Meiller, Blanka Angyal, Paris Wheeler, Candice Davis, and Danelle Stevens-Watkins. Critical consciousness of anti-Black racism: A practical model to prevent and resist racial trauma. *Journal of Counseling Psychology*, 68(1):1–16, January 2021. ISSN 0022-0167. doi: 10.1037/cou0000430. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2020-20397-001&site=ehost-live>. Publisher: American Psychological Association.
- Sean P. Mullen, Neha P. Gothe, and Edward McAuley. Evaluation of the Factor Structure of the Rosenberg Self-Esteem Scale in Older Adults. *Personality and individual differences*, 54(2):153–157, January 2013. ISSN 0191-8869. doi: 10.1016/j.paid.2012.08.009. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3505083/>.
- Hector Najera Catalan. Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study. *Social Indicators Research*, 142, April 2019. doi: 10.1007/s11205-018-1950-z.
- Helen A. Neville, Roderick L. Lilly, Georgia Duran, Richard M. Lee, and LaVonne Browne. Construction and initial validation of the Color-Blind Racial Attitudes Scale (CoBRAS): Journal of Counseling Psychology. *Journal of Counseling Psychology*, 47(1):59–70, January 2000. ISSN 0022-0167. doi: 10.1037/0022-0167.47.1.59. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2000-13659-006&site=ehost-live>. Publisher: American Psychological Association.
- Thomas A. Norton, Hannes Zacher, Stacey L. Parker, and Neal M. Ashkanasy. Bridging the gap between green behavioral intentions and employee green behavior: The role of green psychological climate: Journal of Organizational Behavior. *Journal of Organizational Behavior*, 38(7):996–1015, September 2017. ISSN 0894-3796. doi: 10.1002/job.2178. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2017-01910-001&site=ehost-live>. Publisher: John Wiley & Sons.
- Mike C. Parent. Handling item-level missing data: Simpler is just as good. *The Counseling Psychologist*, 41(4):568–600, May 2013. ISSN 0011-0000, 1552-3861. doi: 10.1177/0011000012445176. URL <http://journals.sagepub.com/doi/10.1177/0011000012445176>.
- Ernest T. Pascarella and Patrick T. Terenzini. Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model. *The Journal of Higher Education*, 51(1):60–75, January 1980. ISSN 0022-1546, 1538-4640. doi: 10.1080/00221546.1980.11780030. URL <https://www.tandfonline.com/doi/full/10.1080/00221546.1980.11780030>.

- Elazar J. Pedhazur and Liora Pedhazur Schmelkin. *Measurement, design, and analysis: an integrated approach*. Lawrence Erlbaum Associates, Hillsdale, N.J., 1991. ISBN 978-0-89859-555-0.
- James A. Pershing and Jana L. Pershing. Ineffective reaction evaluation. *Human Resource Development Quarterly*, 12(1):73–90, 2001. ISSN 1532-1096. doi: 10.1002/1532-1096(200101/02)12:1<73::AID-HRDQ6>3.0.CO;2-7. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/1532-1096%28200101/02%2912%3A1%3C73%3A%3AAID-HRDQ6%3E3.0.CO%3B2-7>.
- Alex L. Pieterse, Katherine Kirkinis, Lynsay Paiko, and Matthew Miller. The Anti-Racism Behavioral Inventory: A Validation Study. *Journal for Social Action in Counseling & Psychology*, 14(2):48–63, 2022. ISSN 2159-8142. doi: 10.33043/JSACP.14.2.48-63. URL <https://openjournals.bsu.edu/jsacp/article/view/3679>. Number: 2.
- Courtney A. Resnick and M. Paz Galupo. Assessing experiences with LGBT microaggressions in the workplace: Development and validation of the Microaggression Experiences at Work Scale: Journal of Homosexuality. *Journal of Homosexuality*, 66(10):1380–1403, 2019. ISSN 0091-8369. doi: 10.1080/00918369.2018.1542207. Publisher: Taylor & Francis.
- William Revelle. An introduction to the psych package: Part II Scale construction and psychometrics. page 97.
- William Revelle and David M. Condon. Reliability from to : A tutorial. Online supplement. *Psychological Assessment*, August 2019a. ISSN 1939-134X, 1040-3590. doi: 10.1037/pas0000754. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/pas0000754>.
- William Revelle and David M. Condon. Reliability from to : A tutorial. *Psychological Assessment*, August 2019b. ISSN 1939-134X, 1040-3590. doi: 10.1037/pas0000754. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/pas0000754>.
- Revelle, William. The Personality Project: An introduction to psychometric theory. URL <http://www.personality-project.org/r/book/#chapter7>.
- Michael T. Roberson and Eric Sundstrom. Questionnaire design, return rates, and response favorableness in an employee attitude questionnaire. *Journal of Applied Psychology*, 75(3):354–357, 1990. ISSN 1939-1854(Electronic),0021-9010(Print). doi: 10.1037/0021-9010.75.3.354.
- Robert G. Green, Katrina D. Murphy, and Shelita M. Snyder. Should demographics be placed at the end or at the beginning of a mailed questionnaires? An empirical answer to a persistent methodological question. *Social Work Research*, 24(4):237–240, December 2000.
- Jenny Rowley. Designing and using research questionnaires. *Management Research Review*, March 2014. ISSN 2040-8269. doi: 10.1108/MRR-02-2013-0027. URL <https://www.emerald.com/insight/content/doi/10.1108/MRR-02-2013-0027/full/html>.
- John Schmid and John M. Leiman. The development of hierarchical factor solutions. *Psychometrika*, 22:53–61, 1957. ISSN 1860-0980. doi: 10.1007/BF02289209. Place: Germany Publisher: Springer.
- Michael F. Schober and Frederick G. Conrad. Survey Interviews and New Communication Technologies. In Frederick G. Conrad and Michael F. Schober, editors, *Envisioning the Survey Interview of the Future*, pages 1–30. John Wiley & Sons, Inc., Hoboken, NJ, USA, November 2007. ISBN 978-0-470-18337-3 978-0-471-78627-6. doi: 10.1002/9780470183373.ch1. URL <http://doi.wiley.com/10.1002/9780470183373.ch1>.

Norbert Schwarz and Daphna Oyserman. Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *AMERICAN JOURNAL OF EVALUATION*, page 34, 2001.

Anneliese Singh. Building a Counseling Psychology of Liberation: The Path Behind Us, Under Us, and Before Us. *The Counseling Psychologist*, 48(8):1109–1130, November 2020. ISSN 0011-0000, 1552-3861. doi: 10.1177/0011000020959007. URL <http://journals.sagepub.com/doi/10.1177/0011000020959007>.

Robert L. Spitzer, Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 166(10):1092, May 2006. ISSN 0003-9926. doi: 10.1001/archinte.166.10.1092. URL <http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.166.10.1092>.

Dawn M. Szymanski and Danielle Bissonette. Perceptions of the LGBTQ College Campus Climate Scale: Development and psychometric evaluation. *Journal of Homosexuality*, 67(10):1412–1428, August 2020. ISSN 0091-8369, 1540-3602. doi: 10.1080/00918369.2019.1591788. URL <https://www.tandfonline.com/doi/full/10.1080/00918369.2019.1591788>.

Susan R. Torres-Harding, Alejandro L. Jr. Andrade, and Crist E. Romero Diaz. The Racial Microaggressions Scale (RMAS): A new scale to measure experiences of racial microaggressions in people of color: Cultural Diversity and Ethnic Minority Psychology. *Cultural Diversity and Ethnic Minority Psychology*, 18(2):153–164, April 2012. ISSN 1099-9809. doi: 10.1037/a0027658. URL <https://ezproxy.spu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=2012-09819-005&site=ehost-live>. Publisher: Educational Publishing Foundation.

Clairice T. Veit and John E. Ware. The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology*, 51(5):730–742, October 1983. ISSN 0022-006X. doi: 10.1037/0022-006X.51.5.730. URL <https://ezproxy.spu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=psyh&AN=1984-02935-001&site=ehost-live>. Publisher: American Psychological Association.