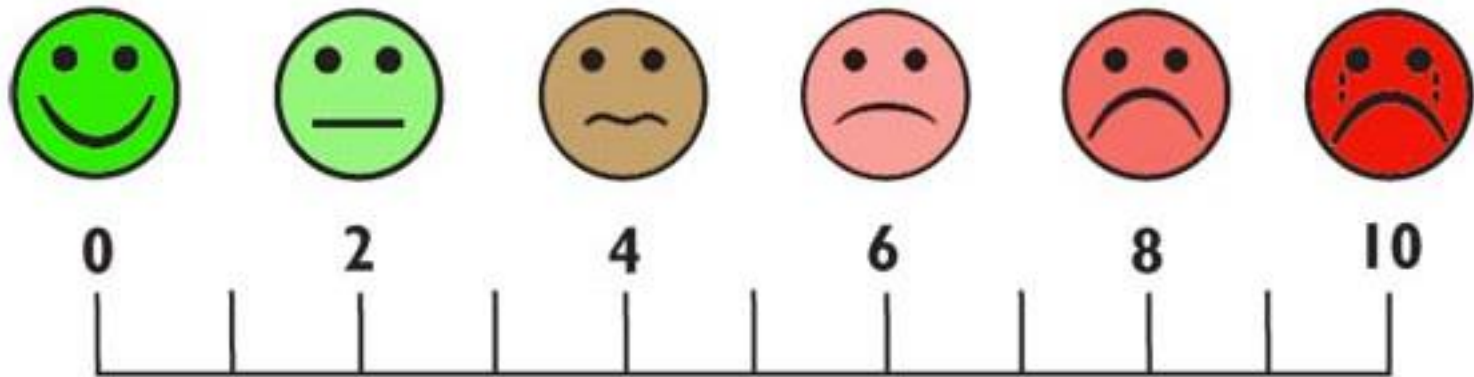




On a Scale of 1 to 10: How do You Feel about Surveys?



Capt Katie Batterton
Capt Kim Hale

Air Education and Training Command
Studies and Analysis Squadron

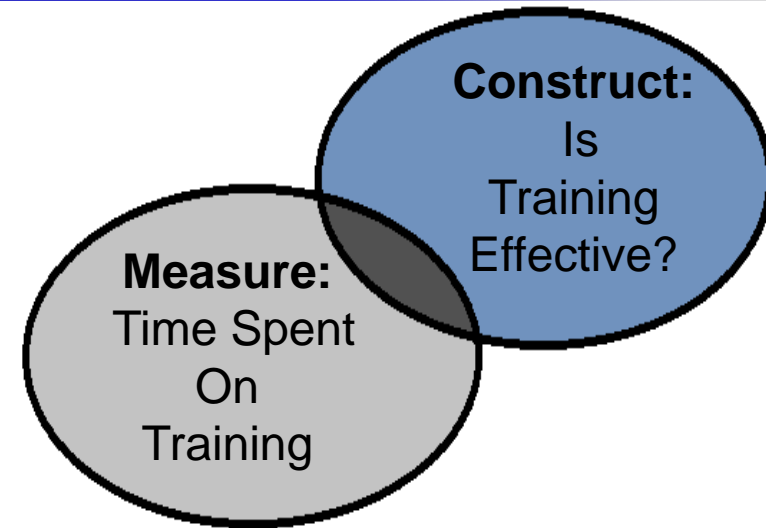
Recruit, Train, and Educate Airmen to Deliver Airpower for America



Survey Design



- Construct – Research Questions
 - What information are we trying to find?
- Measurement – Survey Questions
 - Will answers accurately reflect the construct?
 - What type of questions do we need?
- Target population
 - Who do we want to study?
- Sample design
 - Who are we asking questions?
 - Methods of sample selection
 - Represent key subgroups



Target population



Sample





Question Development



- Question pitfalls to avoid
 - Excessively complex or vague
 - Built in assumptions or perceived intent
- Evaluate question performance
 - Expert review
 - Focus group discussions
 - Pilot test
- Screen responses
 - Nonresponse, atypical answer patterns, illogical answers, etc.
 - Validate using honesty or dummy questions





Question Development



- Question pitfalls to avoid
 - Excessively complex or vague
 - Built in assumptions or perceived intent
- Evaluate question performance
 - Expert review
 - Focus group discussions
 - Pilot test



Example Question:

“Compulsory military training in all countries should be reduced but not eliminated.”



Question Development



- Question pitfalls to avoid
 - Excessively complex or vague
 - Built in assumptions or perceived intent
- Evaluate question performance
 - Expert review
 - Focus group discussions
 - Pilot test



Example Question:

“Compulsory military training in all countries should be reduced but not eliminated.”



attitude

a settled way of thinking or feeling about someone or something, typically one that is reflected in a person's behavior



The Misuse of Likert Scales...

Likert scales: how to (ab)use them

Susan Jamison, 2004

Likert scales reflect ordinal data

Mean & Standard Deviation not appropriate for ordinal data

Becoming common to treat Likert scales as interval data

Nonparametric methods should be used for ordinal data

Likert scales: how to (ab)use them

Susan Jamison

Dipping my toe into the water of educational research, I have recently used Likert-type rating scales to measure student views on various educational interventions. Likert scales are commonly used to measure attitudes, providing a range of responses to a given question or statement.¹ Typically, there are 5 categories of response, from the (example) 1 = strongly disagree to 5 = strongly agree, although there are arguments in favour of scales with 7 or with an even number of response categories.²

between 'strongly disagree' and 'disagree' is equivalent to the amount of feeling between other consecutive categories on the Likert-type scale. The legitimacy of assuming an interval scale for Likert-type categories is an important issue, because the appropriate descriptive and inferential statistics differ for ordinal and interval variables^{3,4} and if the wrong statistical technique is used, the researcher increases the chance of coming to the wrong conclusion about the significance (or otherwise) of his research.

parametric tests require data of interval or ratio level.⁵ Treating ordinal scales as interval scales has long been controversial.

However, these 'values' are commonly ignored by authors, including some who have published in *Medical Education*. For example, the authors of 2 recent papers had used Likert scales but described their data using means and standard deviations and performed parametric analyses such as *t*-tests.^{6,7} This is consistent with Blakie's observation that it has become common practice to assume that Likert-type categories constitute interval-level measurement.⁸ Generally, it is not made clear by authors whether they are aware that some would regard this as illegitimate; no statements are made about an assumption of interval status for Likert data, and no argument scale is supplied.

parametric tests require data of interval or ratio level.⁵ Treating ordinal scales as interval scales has long been controversial.

However, these 'values' are commonly ignored by authors, including some who have published in *Medical Education*. For example, the authors of 2 recent papers had used Likert scales but described their data using means and standard deviations and performed parametric analyses such as *t*-tests.^{6,7} This is consistent with Blakie's observation that it has become common practice to assume that Likert-type categories constitute interval-level measurement.⁸ Generally, it is not made clear by authors whether they are aware that some would regard this as illegitimate; no statements are made about an assumption of interval status for Likert data, and no argument scale is supplied.

It has become common practice to assume that Likert-type categories constitute interval-level measurement.

All of which is very confusing for the novice in pedagogical research. What approach should one take when specialist tests say I think, yet ordinal practice differs?

The response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal, although, as Blakie⁸ points out, "...researchers frequently assume that they are". However, Cohen *et al.*⁹ contend that it is "illegitimate" to infer that the intervals of feeling

Methodological and statistical texts are clear that for ordinal data one should employ the median or mode as the 'measure of central tendency' because the arithmetical manipulations required to calculate the mean (and standard deviation) are inappropriate for ordinal data.¹⁰ "show the numbers generally represent verbal statements. In addition, ordinal data may be described using frequencies (percentages of response in each category). Standard texts also advise that the appropriate inferential statistics for ordinal data are those employing nonparametric tests, such as the sign test, Spearman's *B* test, or the Mann-Whitney *U* test.¹¹ Because

Copyright © 2004 Blackwell Publishing Ltd *Journal of Medical Education* 79(12): 1212-1218
doi:10.1111/j.1365-2006.2004.00101.x

© Blackwell Publishing Ltd *Journal of Medical Education* 2004; 79: 1212-1218

Ten Common Misunderstandings, Misconceptions, Persistent Myths, & Urban Legends about Likert Scales & Likert Response Formats & their Antidotes

James Carifio and Rocco J. Perla, 2007

Response to Susan Jamison's article

Distinguish Likert scales & response formats

Likert scales result in interval data

Likert response items result in ordinal data

Journal of Social Sciences 3 (3): 106-116, 2007
ISSN 1548-0657
© 2007 Science Publications

Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes

James Carifio and Rocco J. Perla
University of Massachusetts Lowell, One University Ave., Lowell, MA 01854
HealthAlliance Hospital, 60 Hospital Road, Lawrence, MA 01845

Abstract: A recent article by Jamison in *Medical Education* outlined some of the (illegally) abuses of "Likert scales" with suggestions about how researchers can overcome some of the (illegally) methodological pitfalls and limitations.¹ However, many of the claims advanced in the Jamison article, as well as a great many of articles it cited, and similar recent articles in medical, health, psychology, and educational journals and books, are themselves common misunderstandings, misconceptions, conceptual errors, persistent myths and "urban legends" about "Likert scales" and their characteristics and qualities that have been propagated and perpetuated across its decades, for a variety of different reasons. This article identifies, analyzes and traces many of these aforementioned problems and presents the arguments, counter arguments and responses that show these many persistent claims and myths about "Likert scales" to be factually incorrect and untrue. Many studies have shown that Likert Scales can respond to single Likert response format items produce interval data and that the *t*-test is very robust to violations of the interval data assumption and moderate skewing and may be used to analyze Likert data (even if it is ordinal, but not as an *ad hoc* means-test "disguise" item, which is simply a current research and analysis practice that must stop. After sixty years, it is more than time to dispel these pervasive research myths and urban legends as well as the various damage and problems they cause, and put them to bed and out of their misery once and for all.

Keywords: formats, Likert, measurement, psychological, scales

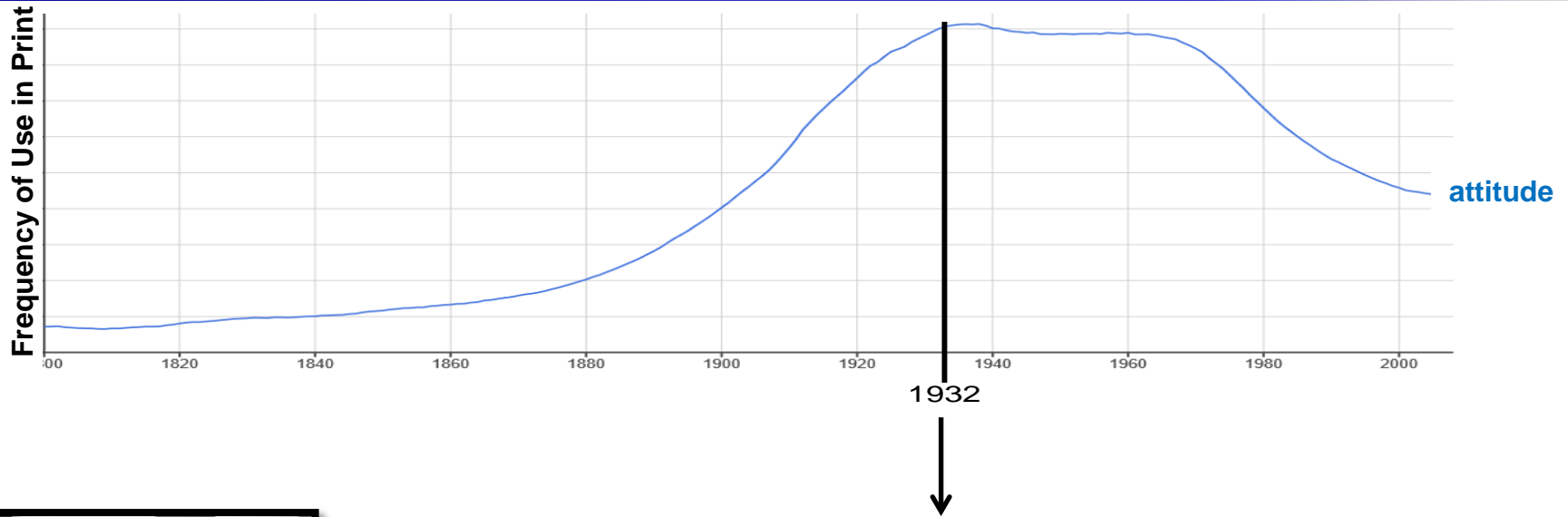
INTRODUCTION

In the process of reviewing literature related to assessments in medical and health education, we came across a recent article by Jamison in *Medical Education* that attempts to outline some of the (illegally) abuses of Likert scales with suggestions of how researchers can overcome some of these methodological pitfalls and limitations.¹ However, many of the claims advanced in the Jamison article relative to Likert "scales," as well as a great many of articles it cited²⁻¹¹ are themselves common misunderstandings, misconceptions, conceptual errors, myths and "urban legends" about Likert scales and their characteristics and qualities that have been propagated and perpetuated across decades, for a variety of different reasons, including a lack of first hand familiarity and understanding of primary sources (i.e., Likert's actual writings), and various and definitive primary empirical studies done by Likert and others (see below). In this respect, Jamison is no different from the dozens of sources over a twenty year period she cites in her article about "Likert scales." Further, this problem is not just confined to the field of medicine and medical education, as the majority of the articles that are the source and progenitors of many of the most important errors and misunderstandings currently exist concerning "Likert scales" are from psychology, education and the field of psychometrics in the fifties and early sixties.¹²⁻¹⁵ These "root of current urban legends" articles, moreover, are additionally more than just "historical context" to anyone who has actually read Likert in the original or contextualized and empirically developed a "Likert scale" according to his theoretical model and writings^{16,17}. This article, therefore, addresses this important problem and a number of persistent misunderstandings, misconceptions, and factual and empirical errors, myths and legends about Likert scales and their characteristics and properties with the hope of helping

Corresponding Author: James Carifio, University of Massachusetts Lowell, USA



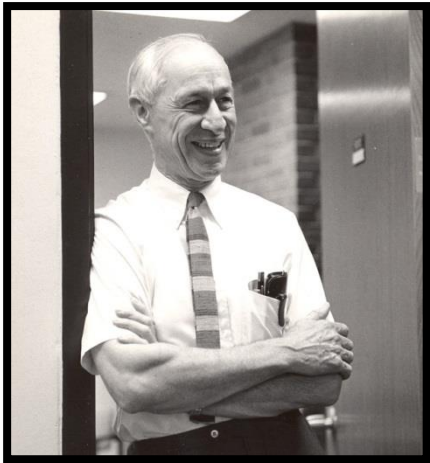
What does Likert *Actually* Say?



A Technique for the Measurement of Attitudes

Rensis Likert, 1932

- Construct an attitude scale from sum of related questions
 - Simplest method
 - As good or better than more complex and existing methods
- 5 point scale more reliable than 3 point scaled questions



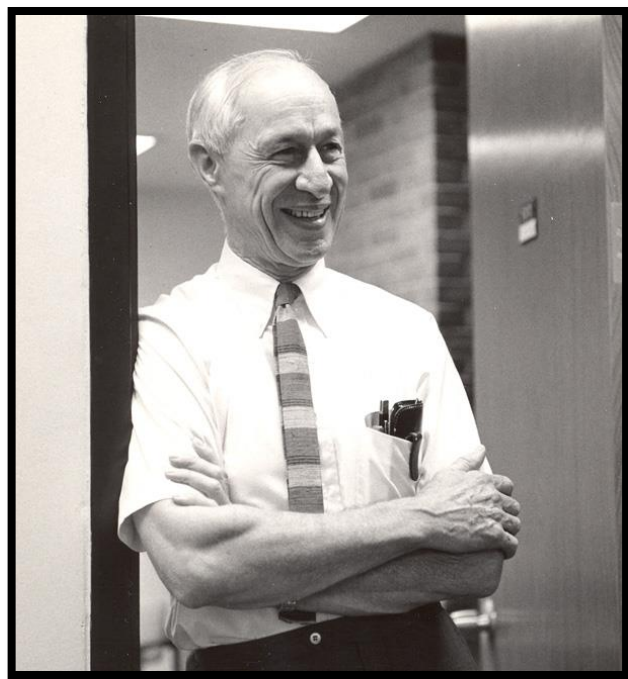
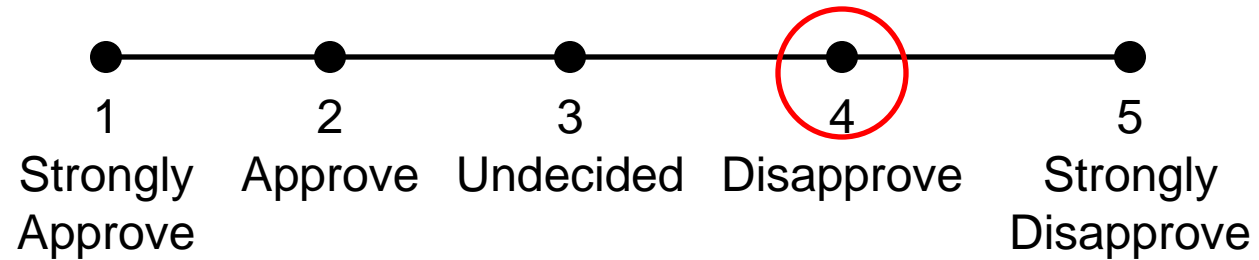
Rensis Likert



What is a Likert Scale?



Likert-Type Response



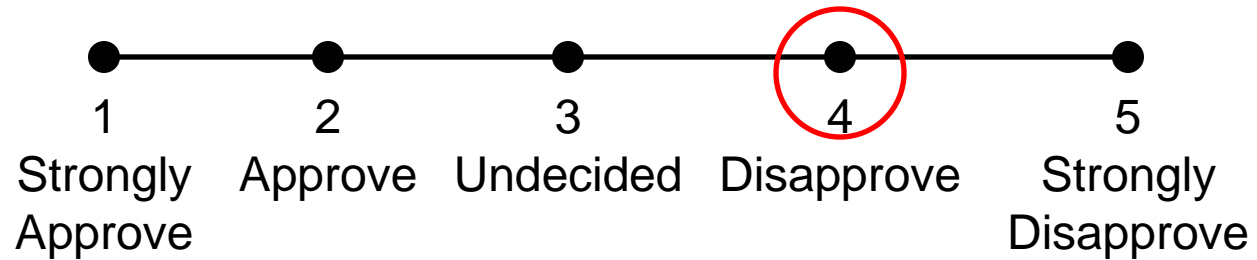
Rensis Likert



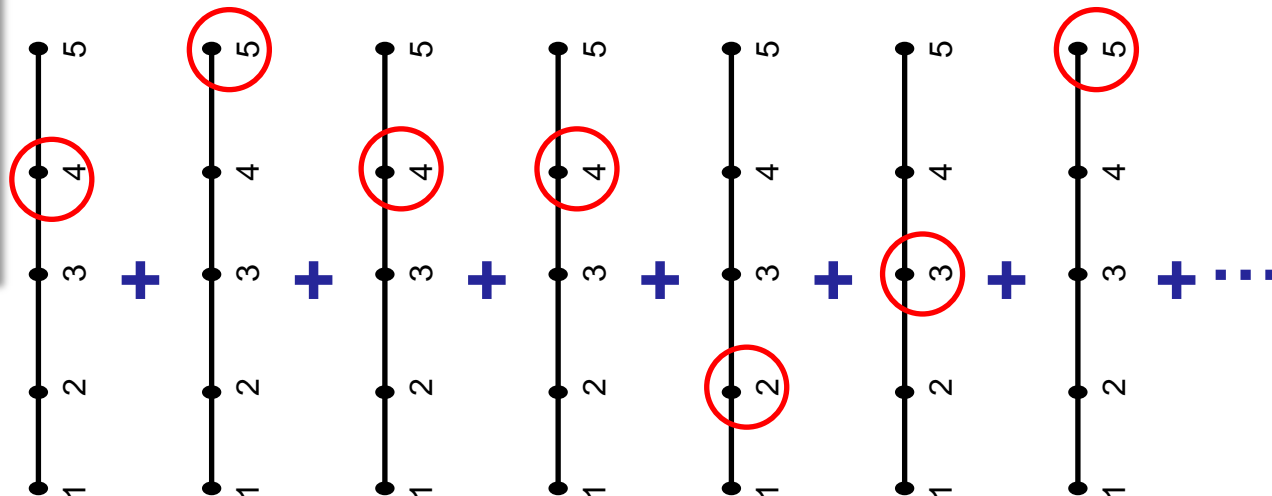
What is a Likert Scale?



Likert-Type Response



Likert Scale



Rensis Likert



How to Construct a Likert Scale

Factor Analysis

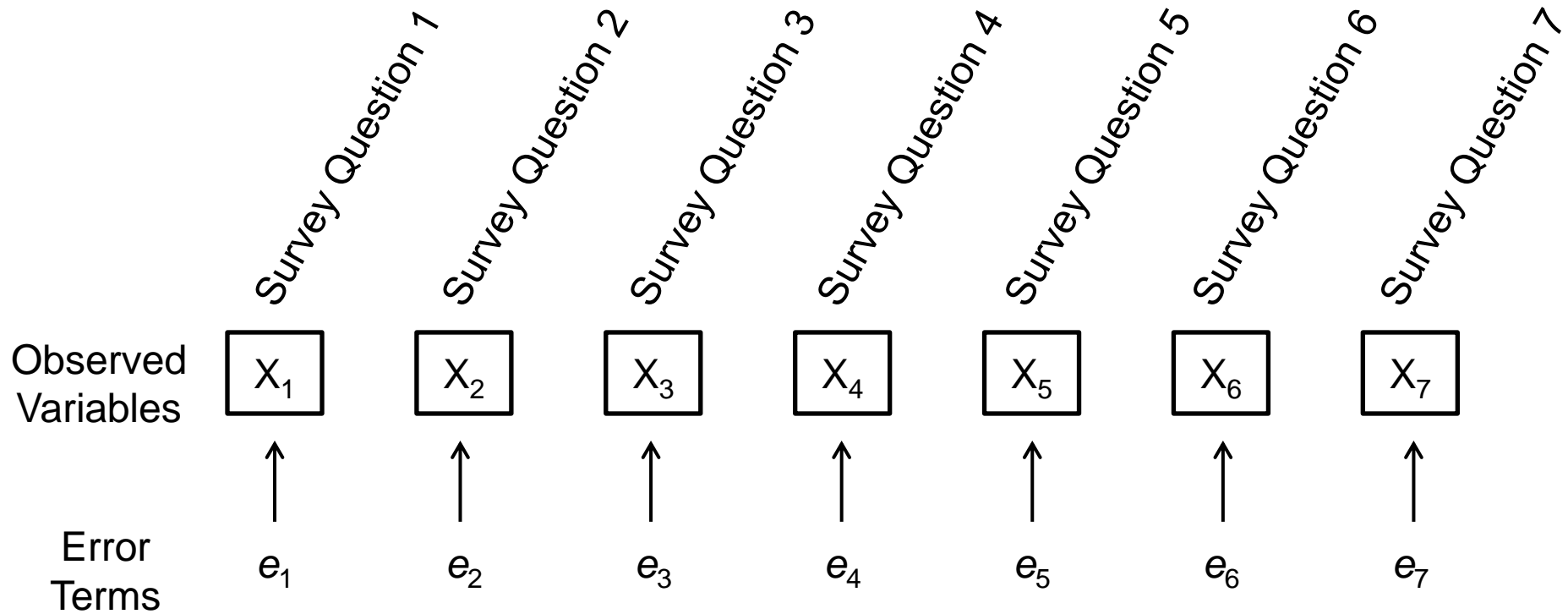


“Factor analysis attempts to simplify complex and diverse relationships that exist among a set of observed variables by uncovering common dimensions or factors that link together the seemingly unrelated variables, and consequently provides insight into the underlying structure of the data.”

-Dillon and Goldstein, 1984



Exploratory Factor Analysis





Exploratory Factor Analysis



Unobservable
Factors

f_1

f_2

Observed
Variables

X_1

X_2

X_3

X_4

X_5

X_6

X_7

Error
Terms



e_1

e_2

e_3

e_4

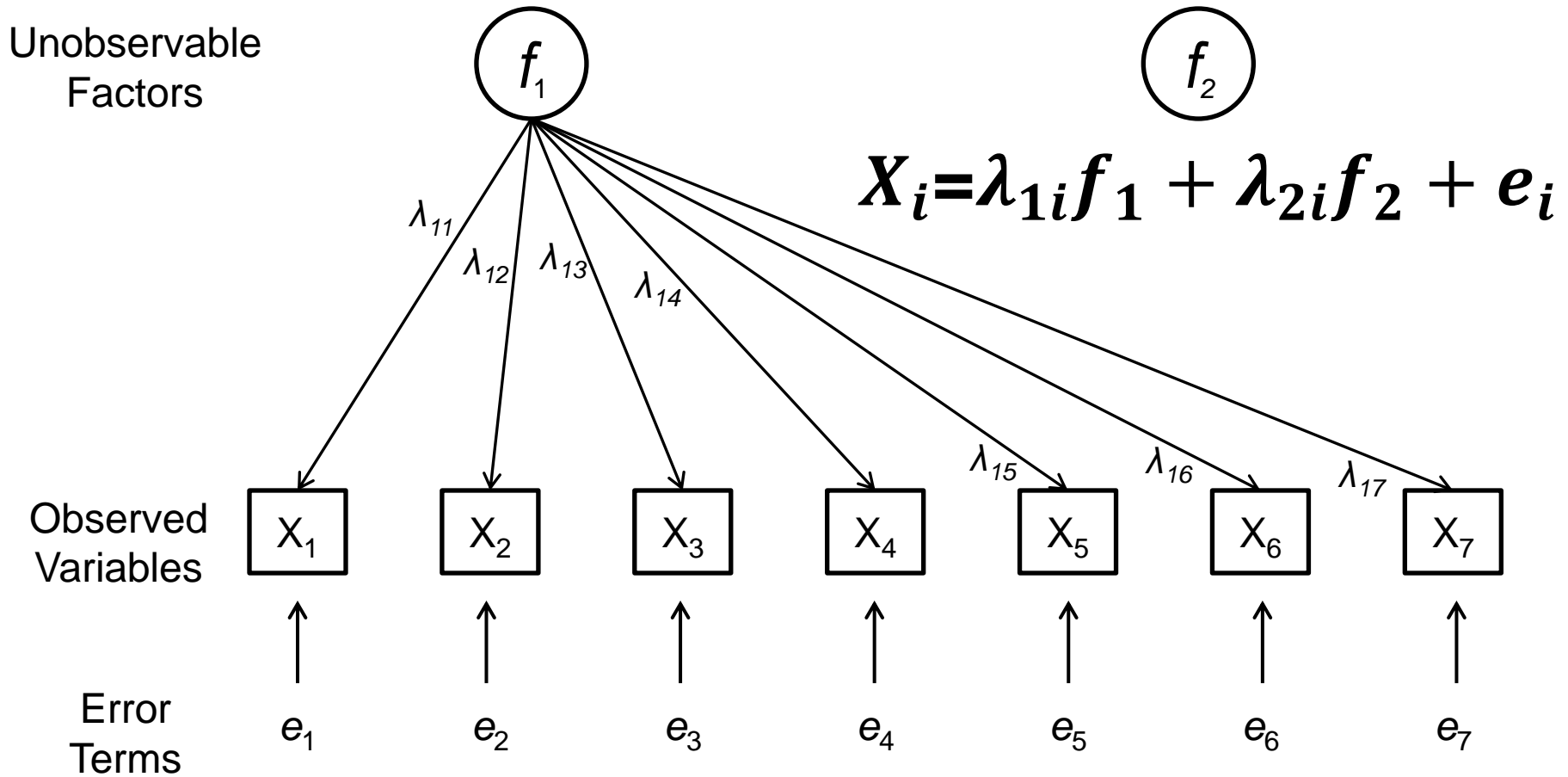
e_5

e_6

e_7

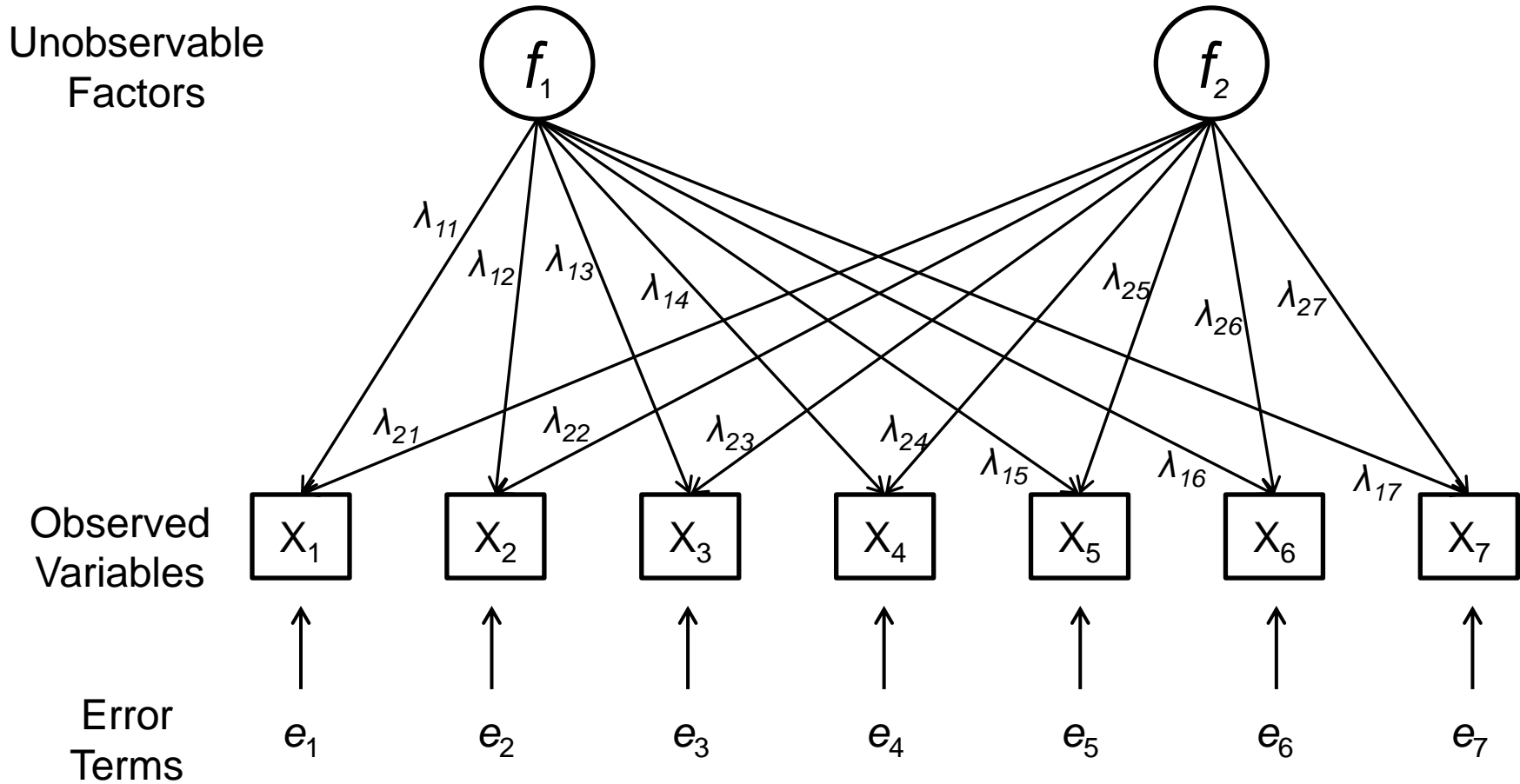


Exploratory Factor Analysis



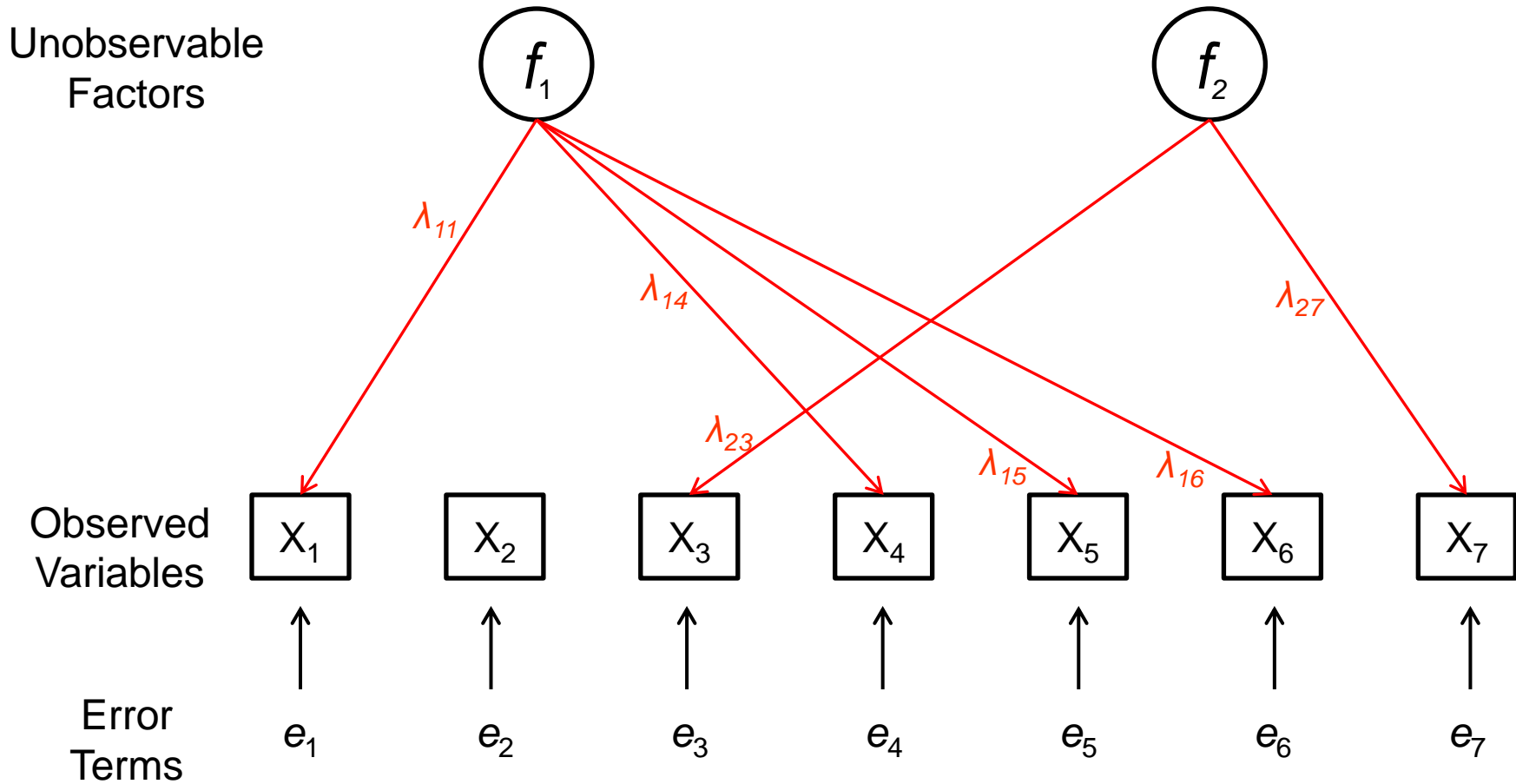


Exploratory Factor Analysis





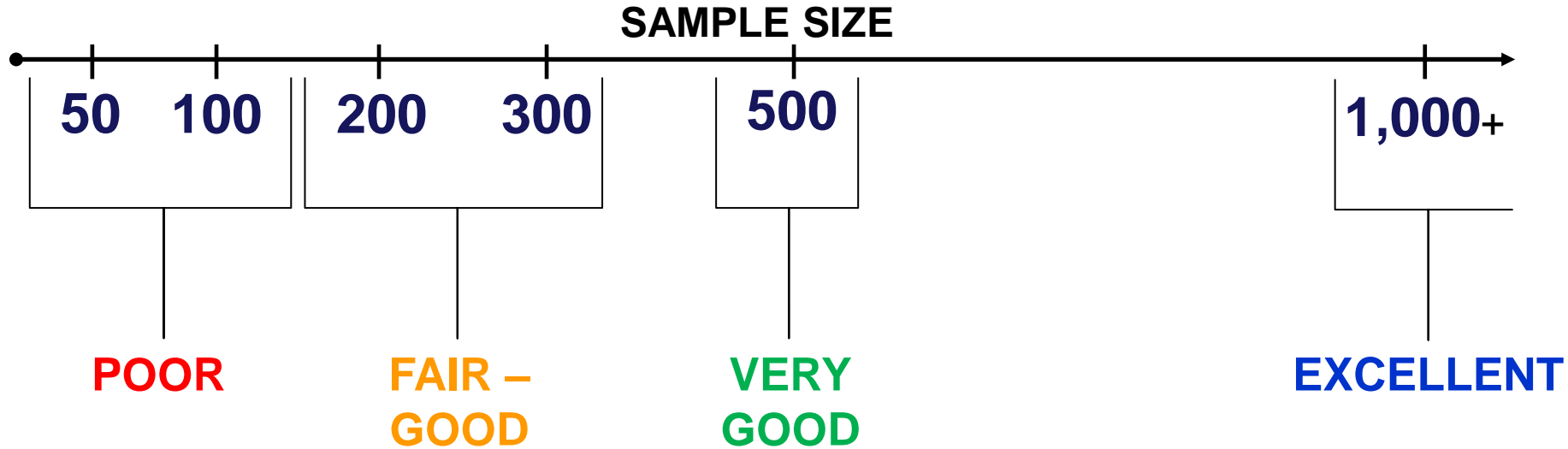
Exploratory Factor Analysis





Sample Size Considerations¹

EFA: the Nuts and Bolts



Strength of factor loadings & number of items also considered

- Several high loading items (>0.8) may require smaller sample size²
- 10:1 ratio of subjects to items is a generally applied rule³

1. Wilson VanVoorhis & Morgan. *Understanding Power and Rules of Thumb for Determining Sample Sizes*. 2007.

2. Guadagnoli & Velicer. *Relation of sample size to the stability of component patterns*. 1988

3. Costello & Osborne. *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis*. 2005

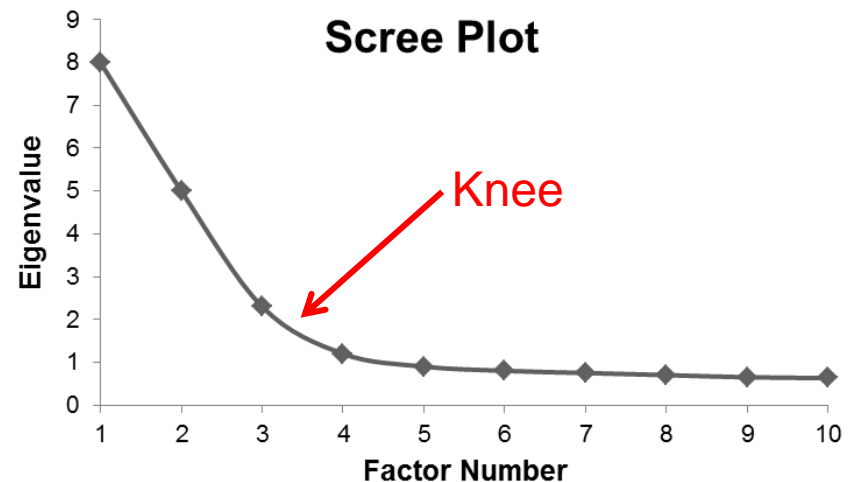


Choosing Factors

EFA: the Nuts and Bolts



- Compute loadings using polychoric correlations¹
- Extract using maximum likelihood or principal factor analysis
- Rotate for interpretation using oblique or orthogonal methods
- Retain based on scree plot



1. Holgado-Tello et al. Polychoric versus Pearson correlations in Exploratory and Confirmatory Factor Analysis of Ordinal Variables. 2010.



Mapping Questions to Factors

EFA: the Nuts and Bolts



- Keep questions with large magnitude loadings
 - 0.4 – 0.7 generally considered low to moderate loadings¹
 - Remove any item with high loadings on multiple factors if possible
- Validate logical relationship between questions in each factor
- Remove questions to increase reliability (Cronbach's Alpha)

$$r_{xx} = \frac{m}{m-1} \left(1 - \frac{\sum_{i=1}^m S_i^2}{S_x^2} \right)$$

- m is number of questions
- S_i^2 is variance of question i for current sample
- S_x^2 is the variance of observed scores

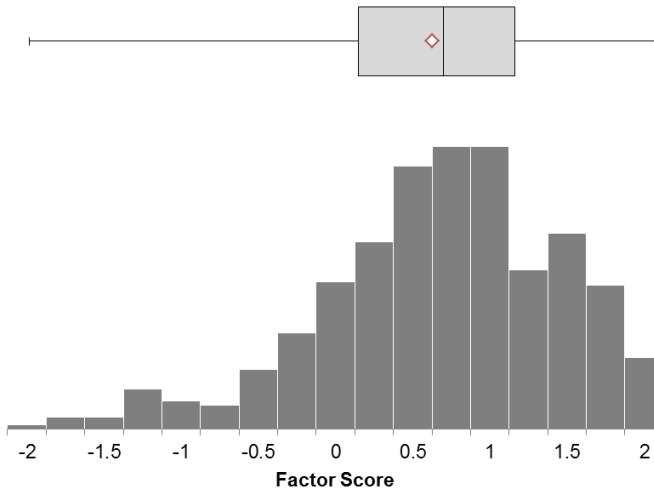
1. Costello & Osborne. *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis*. 2005.



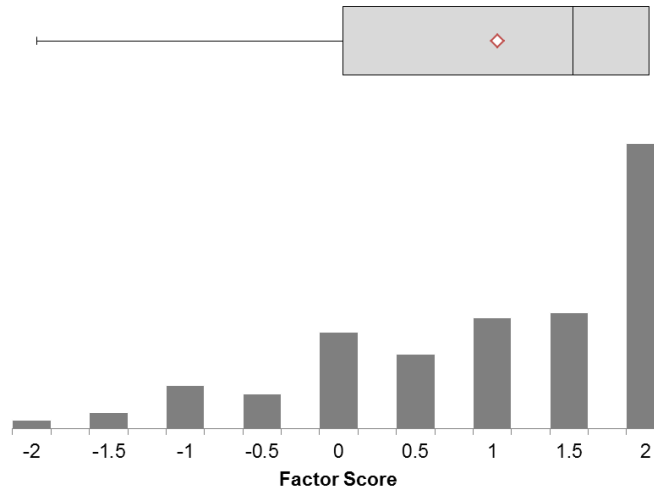
Assessing your Scale¹



- Aim for 4 or more survey questions in each factor²



4 Item Scale



2 Item Scale

- Use Confirmatory Factor Analysis to assess strength
 - Cronbach's Alpha ≥ 0.7
 - Standardized Root Mean Squared Residual (SRMR) ≤ 0.09
 - Tucker-Lewis Index (TLI) ≥ 0.95 (or Bollen's (BL89))
 - Other fit indexes available (see Hu & Bentler, 1999)

1. Hu & Bentler. *Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives*. 1999

2. Costello & Osborne. *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis*. 2005.



AF Survey Example



Survey Question

Factor 1 Factor 2

1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Choose Factors

AF Survey Example



Survey Question

Factor 1 Factor 2

1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Map Questions to Factors

AF Survey Example



Survey Question

Factor 1 Factor 2

1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Validate Logical Relationship

AF Survey Example



Survey Question	Factor 1	Factor 2
1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Validate Logical Relationship

AF Survey Example



Survey Question	<u>Leadership Factor</u>	<u>Factor 1</u>	<u>Factor 2</u>
1. There is cronyism (i.e., good old boys club) within my unit.		-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.		0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.		0.12	0.61
4. I have access to professional development opportunities.		0.29	0.63
5. I am given adequate time to maintain my fitness.		0.29	0.25
6. My leadership encourages and supports professional development.		0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.		0.73	0.31
8. Leaders are appointed based on the “whole person concept.”		0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.		0.63	0.21
10. Leadership properly addresses misconduct.		0.69	0.07
11. My unit has adequate manning to conduct its mission.		0.19	0.27
12. Professional conduct is reinforced in my training.		0.35	0.53
13. I’m satisfied with the leadership preparation training I received.		0.18	0.70



Validate Logical Relationship

AF Survey Example



<u>Survey Question</u>	<u>Factor 1</u>	<u>Factor 2</u>
1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Validate Logical Relationship

AF Survey Example



Survey Question

Training Factor

Factor 1 Factor 2

1. There is cronyism (i.e., good old boys club) within my unit.	-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.	0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.	0.12	0.61
4. I have access to professional development opportunities.	0.29	0.63
5. I am given adequate time to maintain my fitness.	0.29	0.25
6. My leadership encourages and supports professional development.	0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.	0.73	0.31
8. Leaders are appointed based on the “whole person concept.”	0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.	0.63	0.21
10. Leadership properly addresses misconduct.	0.69	0.07
11. My unit has adequate manning to conduct its mission.	0.19	0.27
12. Professional conduct is reinforced in my training.	0.35	0.53
13. I’m satisfied with the leadership preparation training I received.	0.18	0.70



Analyzing a Likert-Scale



Likert-scales should be interval – use parametric methods

<u>Group</u>	<u>Factors</u>				
	Leadership	Training	Organization	Job Requirements	Resources
Junior Enlisted	Positive	Negative	Positive	Very Negative	Negative
Senior Enlisted	Positive	Neutral	Positive	Negative	Negative
Company Grade Officer	Positive	Neutral	Positive	Very Negative	Negative
Field Grade Officer	Positive	Positive	Positive	Negative	Neutral
Junior Civilian	Positive	Neutral	Positive	Neutral	Neutral
Senior Civilian	Positive	Positive	Positive	Neutral	Neutral
Overall	Positive	Negative	Positive	Negative	Negative

Very Negative

Negative

Neutral

Positive

**Likert-scales enhance our ability
to detect differences between groups**



Orphaned Questions

AF Survey Example



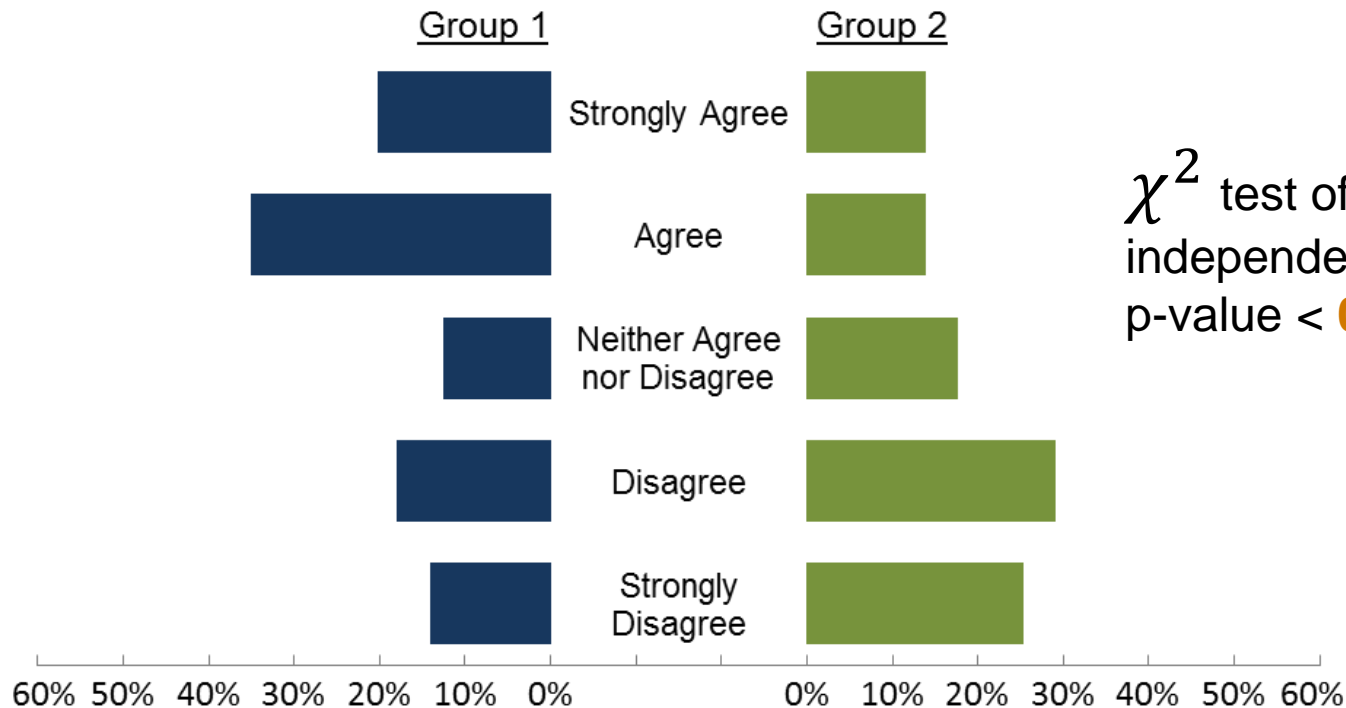
<u>Survey Question</u>	<u>No Factor</u>	<u>Factor 1</u>	<u>Factor 2</u>
1. There is cronyism (i.e., good old boys club) within my unit.		-0.49	-0.19
2. I am comfortable discussing issues with my chain of command.		0.70	0.25
3. Initial skills training adequately prepared me for my primary duties.		0.12	0.61
4. I have access to professional development opportunities.		0.29	0.63
5. I am given adequate time to maintain my fitness.		0.29	0.25
6. My leadership encourages and supports professional development.		0.71	0.31
7. Leadership encourages me to succeed in my Air Force career.		0.73	0.31
8. Leaders are appointed based on the “whole person concept.”		0.52	0.20
9. Decision making is delegated to the appropriate level in my unit.		0.63	0.21
10. Leadership properly addresses misconduct.		0.69	0.07
11. My unit has adequate manning to conduct its mission.		0.19	0.27
12. Professional conduct is reinforced in my training.		0.35	0.53
13. I’m satisfied with the leadership preparation training I received.		0.18	0.70



Analyzing Likert-Type Responses

Likert-type responses are ordinal – use nonparametric methods

5. I am given adequate time to maintain my fitness.

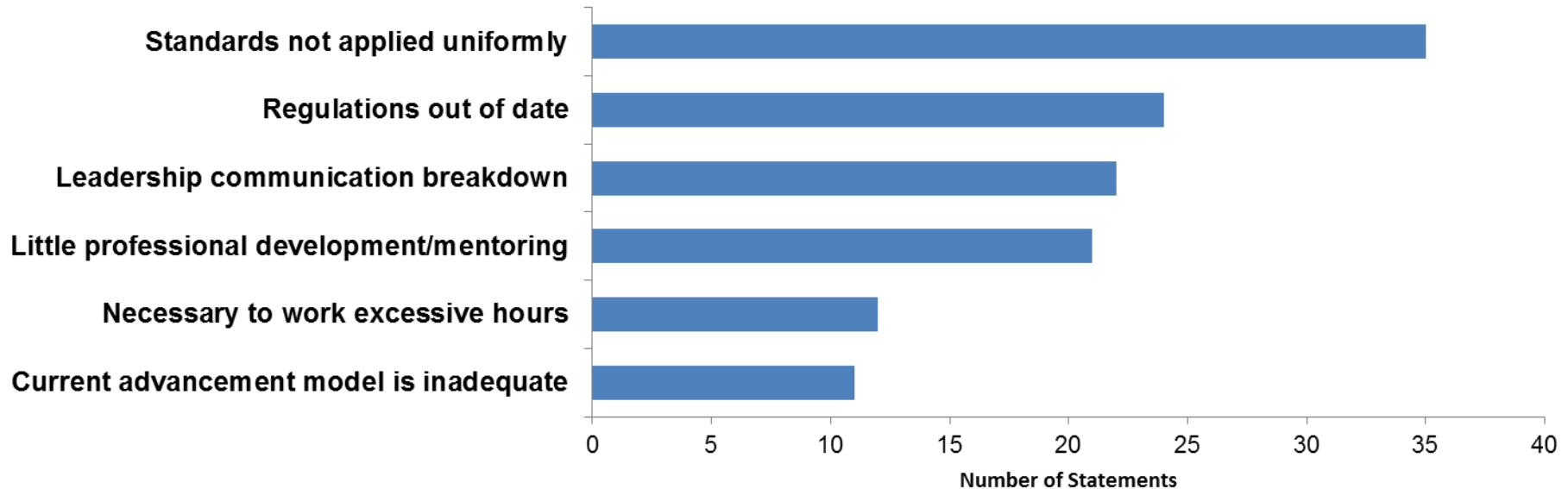


χ^2 test of
independence:
p-value < **0.0001**

Be cautious using single questions to inform decisions



Making Sense of Free Responses





Summary



- Take care when designing a survey
- Use or develop Likert scales – more powerful comparisons
- Likert scales are interval – use parametric methods
- Likert-type responses are ordinal – use nonparametric methods
- Free responses provide an avenue for overall themes and should be included in analysis

