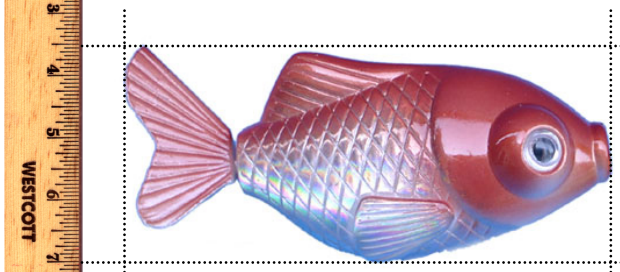
Research methods 10

Measurement of Constructs Validity

Caryn Block
ORLJ 5040
Teachers College
Columbia University

1

The **Goodness** of measures



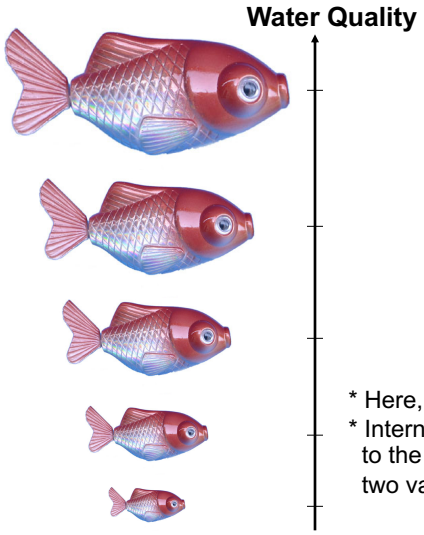
1) Reliability

How precise is our measure?

How much random error is involved in our measurement?

2

The **Goodness** of measures



2) Validity

Do we really measure what we want to measure?

How much systematic error is involved in our measurement?

* Here, validity is related to the quality of a variable
* Internal and external validity are related to the quality of the relationship between two variables.

3

Relationship of **reliability and validity**

- 2 different questions
- Can be measuring something without error (reliably), but it might not be the construct of interest (validity)
- A measure can't be valid if it is not reliable
- Reliability is a necessary, but insufficient condition for validity

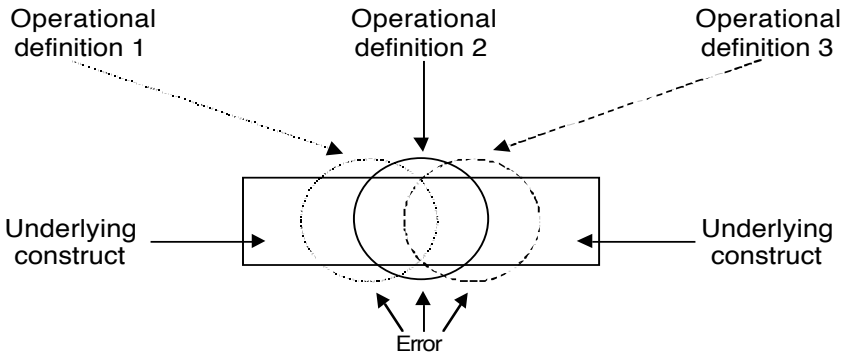
4

Validity

- Measures are high in validity to the extent that they measure what is intended
- Highly reliable measures can measure the wrong construct
- Measures inevitably assess:
  - constructs they are not supposed to assess and
  - only part of the underlying construct of interest
- Multiple measures helpful, especially if they do not share common method variance

5

Operational Definitions Include Irrelevant Components and Fail to Include All Relevant Portions of Underlying Construct

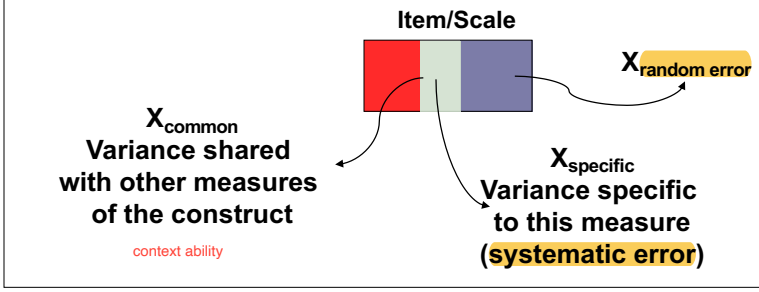


6

Validity of measures

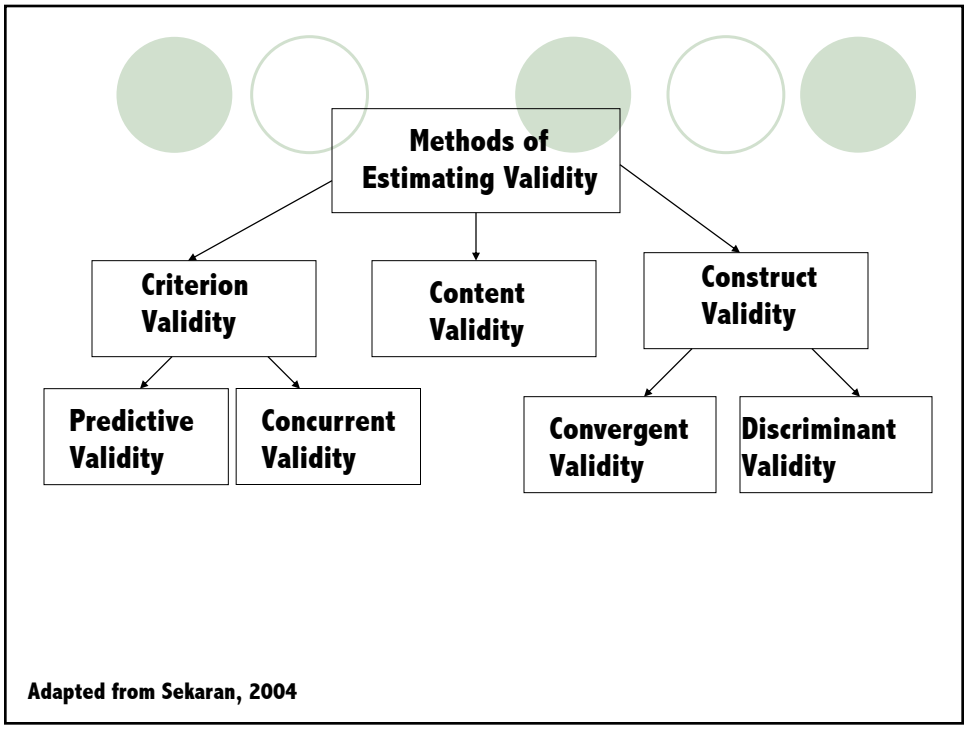
$X_{obs} = X_{true} + X_{random\ error}$

$X_{obs} = X_{common} + X_{specific} + X_{random\ error}$



testing ability like typing spead

7



Adapted from Sekaran, 2004

8

Content Validity

Is the content of this measure representative of the content or the universe of content of the property being measured? (Sampling adequacy)

Content validity is based on judgment. Useful strategy for constructing a measure.

Test with numeric operations → Math abilities

Face validity:

Measure just looks like something it is supposed to measure. Very subjective judgment!

9

Relational Approach to Work Scale Development (Matthews, Buontempo & Block, 2013)

- **Approach to Work:** Preferences for the use of certain types of behaviors and strategies to get work done in organizations
- **Relational Approach to Work:** Emphasizes interdependence, mutual growth and power sharing in work relationships

10

Content Validation: Relational Approach to Work Scale

- **Item generation**
  - Relevant literatures reviewed
  - Items independently generated by 4 researchers
  - Conceptual categories developed
  - 27 possible items
- **Item selection**
  - 223 participants filled out survey
  - Factor analysis
  - Scale reduced to 14 items

- ***Not done:***
  - Subject matter experts rate items for relevance and comprehension
  - CVI computed to retain items on which there is agreement

11

Relational Approach to Work Scale

Table 1 Factor Structure of Approach to Work Scales

(a) Relational, PCA with varimax rotation			
Relational scale items	1	2	3
I try to help out when others are not getting along.	**.719**	.213	
I'll go out of my way to prevent problems that may negatively affect my coworkers.	**.735**	.216	
I offer support when a coworker has a personal problem.	**.857**		
I notice when others are in need of assistance.	**.898**		
I will help someone even if it's not part of my job description.			**.752**
I give my time to help others with work-related problems.			**.821**
I offer to share my expertise with colleagues to contribute to their development.	**.865**		
I like to take time to develop personal relationships at work.	**.731**	.327	
I try to say things to make my coworkers feel good about themselves and their work.		**.592**	.254
I think it's important to foster harmony within my workgroup.	**.863**		
I like to keep my ideas open to the influence of others.	**.720**		
I am willing to do someone else's job if it helps the group.	**.572**		.309
I usually help to organize social events at work.	.216	**.830**	
I often volunteer to do the tasks that help organize the group.	.403	**.639**	

12

Criterion-Related Validity

Does the measure predict an outside criterion?

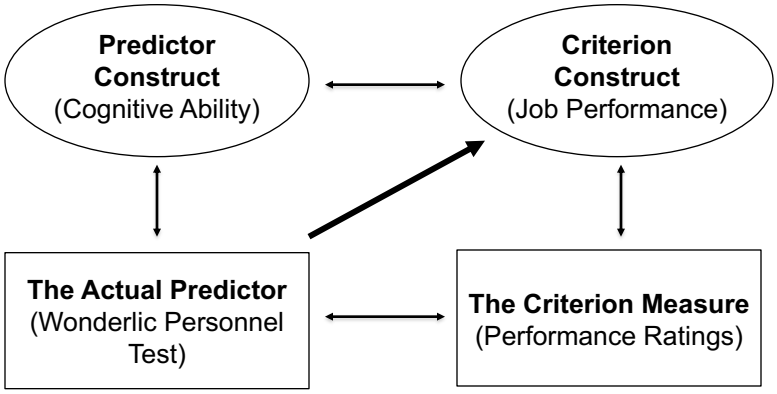
Predictor must be reliable. Can be measured first (predictive validity) or simultaneously (concurrent validity).

Criterion is of primary interest (needs to be reliable and valid measure).

Criterion-related validity is more relevant for solving practical problems (e.g., decision-making, selection processes) than for understanding the measured construct.

13

Criterion-Related Validity



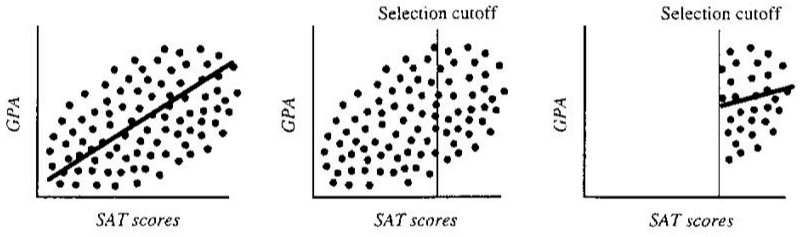
14

Criterion-Related Validity

Type	How	Notes
Predictive (Ideal)	1) Obtain test scores from a group of applicants, but do not use the test in making decisions 2) At some time later, obtain performance measures for those hired/selected, and correlate those measure with test scores	- The key is to have the population in the validity study be similar to the general population of applicants → impractical
Concurrent (Practical alternative)	Obtain test scores and performance measures approximately at the same time, and find correlations between the two	- Practical - Range restriction

15

Restriction of Range



Restriction of range reduces the correlation between test scores and criterion measures

16

Criterion Related Validity: Relational Approach to Work Scale

Table 4 Hierarchical Regression Analysis of Relational Scale

	Dependent: relational scale				
Predictor variables	*B*	*SE*	*t*		$R^2\Delta$
Step 1					
Sex	.29	.11	2.64**		.04
Step 2					
Sex	.07	.11	.59		
Femininity BSRI	.46	.09	5.32	***	.15
Step 3					
Sex	.11	.11	.93		
Femininity BSRI	.46	.09	5.32***		
Masculinity BSRI	.15	.08	1.99*		.02

Note. Unstandardized regression coefficients are reported.
BSRI = Bem Sex-Role Inventory; *$p < .05$. **$p < .01$. ***$p < .00$.

17

Construct validity

- The degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based
  - Can't be demonstrated in a single study
  - Requires an accumulation of evidence

18

Assessing Construct Validity

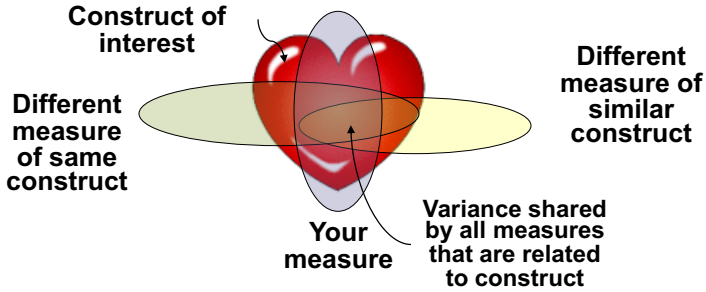
- Evidence for construct validity comes primarily from assessments of:
  - **Convergence**: Your measure is related to other measures of your construct as well as to measures of related constructs
  - **Discriminability**: Your measure is unrelated to measures of your constructs that are unrelated to your construct
- Validity is demonstrated when associations observed with a measure match the theoretical nomological net of the construct

19

Construct validity

Convergence

Different measures of the same construct and of similar constructs are correlated with your measure

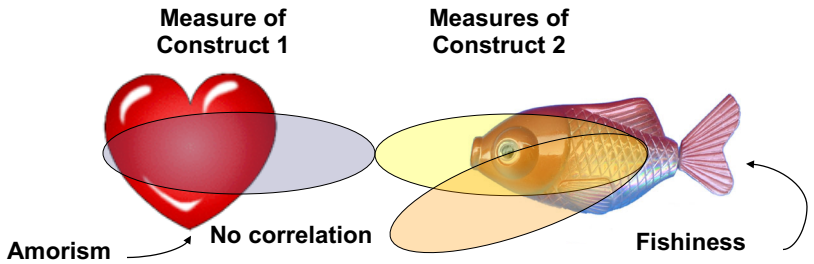


20

Construct validity

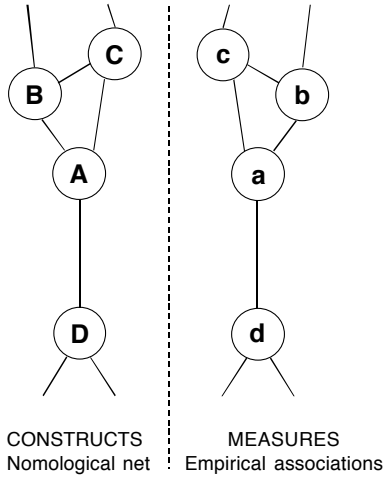
Discriminability

Different measures of an unrelated construct are uncorrelated with your measure



21

Desired Correspondence: Nomological Net and Empirical Associations



22

Construct Validation: Social Dominance Orientation (Pratto, Sidanius, Stalworth & Malle, 1994)

- General attitudinal orientation toward intergroup relations and whether one prefers them to be equal versus hierarchical
- Extent to which one desires that one's in-group dominate and be superior to out-groups

23

Social Dominance Orientation Scale

Items on the 16-Item Social Dominance Orientation Scale

1. Some groups of people are simply inferior to other groups.
2. In getting what you want, it is sometimes necessary to use force against other groups.
3. It's OK if some groups have more of a chance in life than others.
4. To get ahead in life, it is sometimes necessary to step on other groups.
5. If certain groups stayed in their place, we would have fewer problems.
6. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
7. Inferior groups should stay in their place.
8. Sometimes other groups must be kept in their place.
9. It would be good if groups could be equal.
0. Group equality should be our ideal.
11. All groups should be given an equal chance in life.
12. We should do what we can to equalize conditions for different groups.
13. Increased social equality.
14. We would have fewer problems if we treated people more equally.
15. We should strive to make incomes as equal as possible.
16. No one group should dominate in society.

Items 9–16 should be reverse-coded. The response scale was *very negative* (1) to *very positive* (7).

Received July 8, 1993
Revision received May 10, 1994
Accepted May 11, 1994 ■

24

13 Samples

Table 1
Description of Samples

Measure	Sample 1	2	3a	3b	4	5	6	7	8	9	10	11	12	13
Age and gender breakdown														
n	98	463	81	57	190	144	49	224	115	97	231	100	135	46
Age range	17–34	15–56	17–21	17–21		17–35	17–23		17–59	17–36				
% men	50	47		51	47	49	69	50	40	33	54		59	100
% women	50	53		49	53	51	31	50	60	67	46		41	0
Ethnic breakdown														
% Euro-American	48	38		58	38	53	59	49	29	19	67		50	52
% Asian-American	23	40		16	40	24	24	25	51	45	22		33	33
% Hispanic	13	8		4	8	10	15	10	14	17	4		10	11
% Black	15	5		14	5	8	2	6	2	10	4		4	0
% Arab-American	1	2		6	2	0	0	1	3	8	1		1	4
Family income														
Under 20K		12			10	21	6		17	19				
20–30K		9			8	16	6		13	15				
30–40K		11			5	12	8		13	17				
40–55K		17			10	8	10		15	12				
55–70K		20			10	19	10		17	15				
70–100K		14			21	14	19		13	9				
100–150K		8			15	6	19		5	2				
150–200K		5			13	3	11		2	3				
200K+		5			8	1	11		5	6				

Note. Missing numbers indicate that information was not available. Samples 4, 7, 10–13 are probably similar in age distribution and range to Sample 3. Income was self-reported annual family income in thousands of dollars.

25

Predictive Validity: Social Dominance Orientation

- Could SDO predict:
  - Gender
  - Hierarchy role (enhancers, middlers, attenuators)
  - Policy attitudes (e.g. law and order policies, LGBT rights, etc.)
    - Did SDO predict policy attitudes over and above political-economic conservatism?

26

Table 5
Coefficient Alphas of Policy Scales, Correlations With Social Dominance Orientation, and Partial Correlations Controlling for Conservatism, Across Samples

Policy scale	No. of items	Sample 2 (n = 455)	3b (n = 50)	4	5 (n = 129)	6 (n = 37)	8 (n = 100)	9 (n = 89)
				Coefficient α				
Chauvinism	8			.73		.73		.58
Law and order	4	.64	.71		.59	.77[a]	.67	.77[b]
Military programs	3	.67[a]	.75		.73	.67		.59
Gay & lesbian rights	2	.82	.91		.85	.86		.83
Women's rights	4	.63	.72		.69	.63	.80[c]	.74
Social programs	10	.78	.77		.79	.86	.66[c]	.81[d]
Racial policy	7	.71	.81		.68	.72	.60[c]	.77
Miscegeny	2	.96	.97		.93	.94	.91	.87
Environmental policies	5				.71	.80		.76
				Correlations of social dominance orientation and policy items				
Chauvinism					.37**	.49**		.14
Law and order		.08	.23*		.30**	.59**	.24*	.19
Military programs		.33**	.27*		.33**	.70**		.47**
Gay & lesbian rights		−.32**	−.50**		−.29**	−.55**		−.17
Women's rights		−.42**	−.32**		−.39**	−.34*	−.52**	−.42**
Social programs		−.50**	−.31**		−.29**	−.70**	−.55**	−.39**
Racial policy		−.42**	−.46**		−.23**	−.62**	−.54**	−.34**
Miscegeny		−.31**	−.15		−.30**	−.31*	−.25*	−.18
Environmental policies					−.27**	−.40**		−.47**
Republican party preference		.15**	.25*		.24**	.45**	.33**	.27*
				Partial correlations removing political-economic conservatism				
Chauvinism					.40**	.16		.06
Law and order		−.02	.15		.20***	.31*	.25***	.15
Military programs		.16***	.18		.31***	.40**		.46***
Gay & lesbian rights		−.28***	−.32**		−.29***	−.14		−.15
Women's rights		−.38**	−.31**		−.35***	−.27*	−.46***	−.40***
Social programs		−.30***	−.27*		−.30***	−.49**	−.50***	−.37***
Racial policy		−.33***	−.30**		−.22***	−.38**	−.49***	−.31***
Miscegeny		−.28***	−.19		−.31***	−.08	−.23**	−.17*
Environmental policies					−.27***	−.31*		−.46***

[a] Three items. [b] Two items. [c] Six items. [d] Seven items.
* $p < .05$. ** $p < .01$. *** $p < .001$.

27

Discriminant Validity: Social Dominance Orientation

- **Is SDO different than:**
  - Big 5 Personality
  - Self-esteem
  - Dominance

28

Table 6
Coefficient Alphas of Interpersonal Dominance and Self-Esteem and Correlations With Social Dominance Orientation Within Samples

Measure	No. of items	Sample 1 (n = 98)	2 (n = 403)	3a (n = 80)	3b (n = 57)	4 (n = 90)	5 (n = 144)	6 (n = 56)	8 (n = 115)	9 (n = 95)
Coefficient α										
CPI Dominance	35	.82	.79		.79				.71	.71
JPRF Dominance	19	.81	.81		.74				.73	.69
Rosenberg Self-Esteem	10	.87	.87	.88	.88	.90	.87	.85	.84	.83
Correlations										
CPI Dominance		−.11	−.03		−.17				.24**	.20
JPRF Dominance		−.04	.13**		−.17				.01	.04
Rosenberg Self-Esteem		−.09	−.18	.09	.01	.16	−.23**	−.01	−.29**	−.14*

Note. CPI = California Psychological Inventory; JPRF = Jackson Personality Research Form.
* $p < .05$. ** $p < .01$.

29

Convergent Validity: Social Dominance Orientation

- **Is SDO related to:**
  - Empathy
  - Altruism
  - Communality

30

Table 7
Coefficient Alphas of Empathy, Altruism, and PAQ Subscales and Correlations With Social Dominance Orientation

Measure	No. of items	Sample 2 (n = 403)	Sample 3b (n = 57)	Sample 5 (n = 144)	Sample 6 (n = 56)	Sample 8 (n = 115)	Sample 9 (n = 95)
				Coefficient α			
Empathy	28	.76	.72	.77	.75	.75	.73
Concern	7	.73	.77	.66	.75	.69	.68
Distress	7	.71	.70	.67	.74	.61	.52
Perspective-taking	7	.64	.78	.74	.84	.71	.58
Fantasy	7	.71	.79	.70	.72	.70	.64
Altruism	5			.87	.87		
PAQ Communality	23					.76	.78
PAQ Agency	28					.80	.80
				Correlations			
Empathy		−.40**	−.21	−.26**	−.36*	−.38**	−.24*
Concern		−.45**	−.51**	−.47**	−.41**	−.53**	−.40**
Distress		−.03	−.11	.10	−.16	.22**	.21*
Perspective-taking		−.30**	.05	−.20*	−.16	−.39**	−.15
Fantasy		−.23**	.01	−.06	−.25*	−.21*	−.23*
Altruism				−.32**	−.24*		
PAQ Communality						−.42**	−.24*
PAQ Agency						−.10	−.08

Note. PAQ = Personal Attributes Questionnaire.
* $p < .05$. ** $p < .01$.

31

In sum…

Table 9
Average Correlations and Significance Tests Across Samples Between Social Dominance Orientation and Personality Variables, Ideologies, and Policy Attitudes

Measure	Mean r	n	Z
	Personality variables		
Concern for others	−.46	6	−8.92
Communality	−.33	2	−4.84
Tolerance	−.30	3	−5.31
Altruism	−.28	2	−3.98
	Ideologies		
Anti-Black racism	.55	6	15.05
Noblesse oblige	−.57	10	20.30
Nationalism	.54	8	15.96
Sexism	.47	12	14.91
Equal opportunities	.46	3	7.51
Patriotism	.45	3	6.84
Cultural elitism	.40	3	6.94
Political-economic conservatism	.38	8	10.26
Just World	.27	2	3.58
Protestant Work Ethic	.11	3	1.25
	Policy attitudes		
Social programs	−.47	6	−12.74
Racial policy	−.44	6	−11.74
Women's rights	−.40	6	−11.52
Military programs	.44	5	−10.12
Gay & lesbian rights	−.37	5	−8.79
Environmental programs	−.38	3	−6.16
Chauvinism	.34	3	5.34
Miscegeny	−.25	6	−7.36
Republican party preference	.28	6	7.08
Law and order	.28	6	6.38

Note. All Zs were significant at $p < .0001$ except for Just World (p = .0002) and Protestant Work Ethic (p = .10). The mean r was computed using Fisher's z; n denotes number of samples.

32

Construct Validity: Multitrait–Multimethod (MTMM) Matrix

- Table of correlation coefficients for multiple constructs measured using multiple methods
- Allows researchers to evaluate convergent and discriminant validity
  - **Reliability coefficients** (monotrait–monomethod)
  - **Convergent validity coefficients** (monotrait–heteromethod)
  - **Discriminant validity coefficients** (heterotrait–monomethod)
  - **Nonsense coefficients** (heterotrait–heteromethod)

33

MTMM Matrix: Correlations between Attitudes toward Women and Attitudes toward Men

	Paper-and-Pencil Questionnaire		Observations of Behavior	
	ATW	ATM	ATW	ATM
Questionnaire				
ATW	(.90)			
ATM	.30	(.90)		
Behavior				
ATW	.70	.10	(.90)	
ATM	.10	.70	.30	(.90)

Note: The coefficients in parentheses are reliability estimates.

34

MTMM Matrix Indicating Lack of Convergent and Discriminant Validity

	Paper-and-Pencil Questionnaire		Observations of Behavior	
	ATW	ATM	ATW	ATM
Questionnaire				
ATW	(.90)			
ATM	.80	(.90)		
Behavior				
ATW	.40	.30	(.90)	
ATM	.30	.40	.80	(.90)

Note: The coefficients in parentheses are reliability estimates.

35

Construct validity

- **Multitrait Multimethod Matrix**
  - A table of correlations that establish both convergent and discriminant validity of a measure
  - **Convergent validity** established when
    - Validity diagonals > 0
  - **Discriminant validity** established when
    - Validity diagonal > heterotrait heteromethod triangle
    - Validity diagonal > heterotrait monomethod triangle
  - **Reliability** established when
    - Reliability diagonals sufficiently high

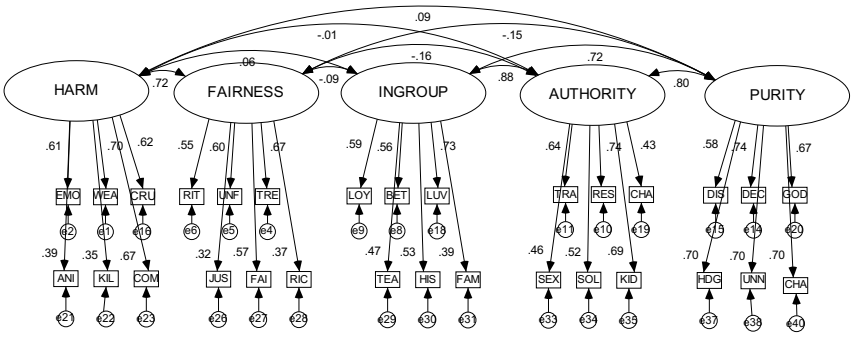
36

Construct Validity: Factor Analysis

- Does measure tap single or multiple dimensions?
- Do dimensions reflect construct(s) in expected ways?
- Assumes that responses to items result from smaller number of latent constructs
- Requires many participants
- Exploratory (EFA)
  - Used when not possible to predict number and nature of dimensions
  - Interpretations of EFA results often subjective
- Confirmatory (CFA)
  - Used when researcher has prediction about number and nature of dimensions or when competing predictions are limited
  - Yields single unique solution and provides significance tests
  - Provides general framework for addressing convergent validity, discriminant validity, and measurement error

37

Moral Foundations Questionnaire CFA (Graham et al., 2011). Reproduced with permission of APA.



38

Summary: Reliability vs. Validity

Reliability	Validity
• How precise is our measure? • A reliable measure can be invalid. • Involves convergence of similar measures (on the same trait). • Is related to random error	• Do we really measure what we want to measure? • A valid measure cannot be unreliable. • Involves convergence of different measures (on the same trait). • Is related to systematic error

39

Cohen (1990)
Things I Have Learned (So Far)

- **Less is more**
  - Except for N
  - Variables in a study
  - Reporting numerical results
- **Simple is better**
  - Reporting of data – graphing, scatterplots
  - Unit weighting

40

Cohen (1990)

Is p<.05 really a magic number?

- **Power analysis** – the cost of overvaluing Type I error
  - **alpha** (Type I error)
    - Probability of falsely rejecting null
    - Concluding there is an effect when there isn't
    - Typically held constant at .05
  - **N** (sample size)
  - **ES** (effect size)
    - Low, moderate or strong effect in the literature?
    - Can be correlation or d= X_1-X_2/sd
  - **beta** (Type II error)
    - Probability of falsely accepting null
    - Concluding there is no effect when there is one
    - Typically very high in most studies (.5-.8)
- Recognize both Type 1 & Type II error in your own research and make an informed decision

41