Validity: Empirical Estimates

PSYC3302: Psychological Measurement and Its Applications

Mark Hurlstone
Univeristy of Western Australia

Week 6

Learning Objectives

- Nomological Network
- Methods of evaluating construct validity
  1. Focussed associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity
- Factors affecting validity coefficients
  1. True associations
  2. Measurement error
  3. Restricted range
- Guidelines for interpreting validity coefficients

Nomological Networks

- Recall that convergent and discriminant evidence reflects the degree to which test scores have the correct pattern of associations with other variables
- The conceptual foundation of a construct includes the connections between the construct and a variety of other psychological constructs
- The interconnections between a construct and other related constructs are known collectively as a **nomological network**

Nomological Networks: Example

- Baumeister and Leary (1995) introduced the construct "need to belong"
- They defined this as "the drive to form and maintain at least a minimum quantity of lasting, positive, and significant interpersonal relationships"
- Leary et al. (2006) theorised about the nomological network surrounding this construct and noted that:
  - the need to belong is similar to constructs such as the *need for affiliation*, the *need for intimacy*, *sociability*, and *extraversion*, but
  - unrelated to *conscientiousness*, *openness to experience*, and *self-esteem*

Nomological Networks

- Leary et al. (2006) proposed the following nomological network for the "need to belong" construct

Table: Nomological network for the construct "need to belong".

Positive	***Negative***	***Non-correlated***
Need for affiliation	Social isolation	Conscientiousness
Sociability		Openness to experience
Extraversion		Self-esteem

- A critical part of the validation process is estimating the degree to which test scores actually show the predicted pattern of associations

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Focused Associations

- There are cases where particular correlations between test scores and certain variables are "make-or-break"
- Consider the correlation between Scholastic Achievement Test (SAT) scores and first-year university marks
- The SAT is a standardized test designed to help universities select students
- It is very similar to an intelligence test, although it is somewhat more focussed on crystallised intelligence (e.g., vocabulary)
- For the SAT to be interpreted as a valid indicator of university performance, it must actually correlate with university marks

Focused Associations

- Empirical research has yielded a correlation of .55 between SAT scores and first year university grades
- Another term for this particular focussed association is predictive validity—a type of criterion validity discussed in last week's lecture
- A more general term is **validity coefficient**
- If research reveals that a test's validity coefficients are generally large, then we have more confidence in using the test for its intended purpose

Validity Generalization

- The SAT and university grade correlation has been estimated based on 110,000 students from more than 25 universities
- Ideally, test users like to see validity coefficients that are generalizable to a broad array of people and circumstances
- Validity generalization is a process of evaluating a test's validity coefficients across a large set of studies

Validity Generalization

- In practice, many measures rely upon a small number of validity studies which include fewer than 400 participants
- Thus, in a lot of cases, it is an assumption that the test scores would be valid in a scenario different to that where it was tested
- For example, a measure of leadership may be useful for senior managers in banking, but it might not be useful for managers in the construction industry
- Perhaps different factors define leadership success in the construction industry, in comparison to the banking industry
- Validity generalization studies are intended to evaluate the predictive utility of a test's scores across a range of settings, times, situations, etc

Validity Generalization

- The textbook mentions the example of 25 studies which examined the association between conscientiousness and job performance
- Different results might be expected to be revealed, because they are based on different types of jobs
  - accountants, lecturers, sales people
- But some of the variance in the validity coefficients may be due to the manner in which job performance was measured
  - in some cases, job performance may be measured by amount of revenue generation
  - in another case, it might be measured based on peer and/or manager ratings

Validity Generalization

- Validity generalization studies can essentially address three questions:
  1. estimate the average level of predictive validity across studies
  2. estimate the degree of variability associated with the validity coefficients
  3. identify sources of systematic variability in the validity coefficients

Validity Generalization

- Validity generalization studies can essentially address three questions:
  1. estimate the average level of predictive validity across studies
  2. estimate the degree of variability associated with the validity coefficients
  3. identify sources of systematic variability in the validity coefficients

Validity Generalization

- Validity generalization studies can essentially address three questions:
  1. estimate the average level of predictive validity across studies
  2. estimate the degree of variability associated with the validity coefficients
  3. identify sources of systematic variability in the validity coefficients

Validity Generalization

- Validity generalization studies can essentially address three questions:
  1. estimate the average level of predictive validity across studies
  2. estimate the degree of variability associated with the validity coefficients
  3. identify sources of systematic variability in the validity coefficients

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Sets of Correlations

- The nomological network surrounding a construct does not always focus on a small number of pertinent criterion variables
- Sometimes a construct's nomological network incorporates a wide variety of other constructs, with differing levels of association with the main construct
- In such cases, researchers must compute the correlations between the test of interest and measures of many criterion variables
- They will then "eyeball" the correlations
- A subjective judgement is then made about the degree to which the pattern of coefficients matches that expected

Example: Perfectionism

- The Perfectionism Inventory (PI; Hill et al., 2004) was designed to measure eights facets of perfectionism (e.g., concern over mistakes, organization, planfulness)
- The authors administered their inventory, in addition to:
  - Multidimensional Perfectionism Scale
  - Multidimensional Perfectionism Scale
  - Fear of Negative Evaluation Scale
  - Brief Symptom Inventory (Depression, Anxiety, OCD)
  - Obsessive Compulsive Inventory
  - Marlowe-Crowne Social Desirability Scale
- Thus, they could evaluate convergent and divergent validity based on the pattern of correlations

Example: Perfectionism

Correlations Between Perfectionism Inventory Scales and Related Measures

Scale	*CM*	*HS*	*NA*	*OR*	*PP*	*PL*	*RU*	*SE*	*CP*	*SEP*	*PL–C*
Perfectionism: MPS–F[a]											
Concern Over Mistakes	.82	.43	.58	.18	.38	.30	.70	.52	.47	.78	.72
Doubts About Actions	.63	.37	.60	.24	.20	.38	.70	.43	.47	.67	.65
Parental Criticism	.41	.25	.20	–.03^{ns}	.60	.02^{ns}	.32	.17	.14	.49	.36
Parental Expectations	.31	.27	.18	.07^{ns}	.85	.06^{ns}	.29	.32	.23	.53	.43
Personal Standards	.47	.50	.36	.45	.39	.44	.52	.72	.70	.55	.71
Organization	.12	.36	.18	.89	.11**	.49	.31	.51	.76	.23	.55
Perfectionism: MPS–HF[b]											
Self-Oriented	.47	.42	.34	.47	.42	.45	.55	.79	.71	.57	.73
Other-Oriented	.33	.62	.14**	.29	.30	.26	.37	.42	.53	.36	.51
Socially-Prescribed	.65	.35	.49	.16**	.58	.21	.61	.42	.38	.74	.65
Symptoms: BSI[c]											
Somatic Complaints	.35	.14*	.31	.13*	.11*	.13*	.34	.17	.19	.35	.31
Depression	.46	.16**	.46	.03^{ns}	.15**	.18	.46	.13*	.17	.49	.39
Obsessive–Compulsive	.40	.14**	.46	.08^{ns}	.10*	.19	.46	.18	.19	.45	.37
Anxiety	.42	.28	.42	.22	.25	.25	.49	.29	.35	.50	.49
Interpersonal Sensitivity	.52	.18	.68	.17	.13*	.22	.56	.27	.28	.60	.51
Hostility	.41	.30	.31	.10*	.21	.05^{ns}	.39	.15**	.20	.42	.36
Phobic Anxiety	.39	.14**	.39	.13*	.15**	.13*	.39	.15**	.21	.42	.37
Paranoia	.48	.28	.49	.18	.21	.21	.54	.30	.33	.55	.51
Psychoticism	.49	.19	.48	.09^{ns}	.16**	.19	.49	.17	.22	.51	.43
Global Severity Index	.54	.24	.55	.16	.20	.21	.57	.25	.29	.59	.51
Obsessive–Compulsive Inventory[d]											
Frequency	.43	.24	.45	.39	.08^{ns}	.34	.52	.42	.47	.47	.54
Distress	.50	.28	.49	.40	.03^{ns}	.33	.60	.44	.48	.51	.57
Fear of Negative Evaluation[a]	.63	.26	.83	.16	.20	.31	.64	.33	.34	.73	.62
Social Desirability: MCSDS[c]	–.15**	–.17	–.09*	–.04^{ns}	–.14**	–.09*	–.18	–.16	–.12**	–.18	–.18

Notes. For all correlations, $p < .001$ (except as noted). CM = Concern Over Mistakes; HS = High Standards for Others; NA = Need for Approval; OR = Organization; PP = Perceived Parental Pressure; PL = Planfulness; RU = Rumination; SE = Striving for Excellence; CP = Conscientious Perfectionism; SEP = Self-Evaluative Perfectionism; PI–C = Perfectionism Inventory Composite score; MPS–F = Frost's Multidimensional Perfectionism Scale; MPS–HF = Hewitt and Flett's Multidimensional Perfectionism Scale; BSI = Brief Symptom Index; MCSDS = Marlowe–Crowne Social Desirability Scale.

[a]$n = 613$. [b]$n = 355$. [c]$n = 368$. [d]$n = 207$.

*$p < .05$, one-tailed. **$p < .01$, one-tailed. $^{ns}p > .05$, all one-tailed.

Example: Perfectionism

- Examination of the pattern of correlations revealed convergent and discriminate evidence for the PI:
  - Concern Over Mistakes scale of the PI strongly associated with a Concern Over Mistakes Scale from a different measure of perfectionism
  - Striving For Excellence scale of the PI strongly associated with Personal Standards scale and a Self-Oriented Perfectionism scale
  - Rumination, Concern Over Mistakes, and Need For Approval scales of PI strongly associated with fear of negative evaluation, frequency, and severity of symptoms of obsessive-compulsive disorder

Example: Perfectionism

- Examination of the pattern of correlations revealed convergent and discriminate evidence for the PI:
  - Concern Over Mistakes scale of the PI strongly associated with a Concern Over Mistakes Scale from a different measure of perfectionism
  - Striving For Excellence scale of the PI strongly associated with Personal Standards scale and a Self-Oriented Perfectionism scale
  - Rumination, Concern Over Mistakes, and Need For Approval scales of PI strongly associated with fear of negative evaluation, frequency, and severity of symptoms of obsessive-compulsive disorder

Example: Perfectionism

Correlations Between Perfectionism Inventory Scales and Related Measures

Scale	*CM*	*HS*	*NA*	*OR*	*PP*	*PL*	*RU*	*SE*	*CP*	*SEP*	*PL–C*
Perfectionism: MPS–F[a]											
Concern Over Mistakes	.82	.43	.58	.18	.38	.30	.70	.52	.47	.78	.72
Doubts About Actions	.63	.37	.60	.24	.20	.38	.70	.43	.47	.67	.65
Parental Criticism	.41	.25	.20	–.03^{ns}	.60	.02^{ns}	.32	.17	.14	.49	.36
Parental Expectations	.31	.27	.18	.07^{ns}	.85	.06^{ns}	.29	.32	.23	.53	.43
Personal Standards	.47	.50	.36	.45	.39	.44	.52	.72	.70	.55	.71
Organization	.12	.36	.18	.89	.11**	.49	.31	.51	.76	.23	.55
Perfectionism: MPS–HF[b]											
Self-Oriented	.47	.42	.34	.47	.42	.45	.55	.79	.71	.57	.73
Other-Oriented	.33	.62	.14**	.29	.30	.26	.37	.42	.53	.36	.51
Socially-Prescribed	.65	.35	.49	.16**	.58	.21	.61	.42	.38	.74	.65
Symptoms: BSI[c]											
Somatic Complaints	.35	.14*	.31	.13*	.11*	.13*	.34	.17	.19	.35	.31
Depression	.46	.16**	.46	.03^{ns}	.15**	.18	.46	.13*	.17	.49	.39
Obsessive–Compulsive	.40	.14**	.46	.08^{ns}	.10*	.19	.46	.18	.19	.45	.37
Anxiety	.42	.28	.42	.22	.25	.25	.49	.29	.35	.50	.49
Interpersonal Sensitivity	.52	.18	.68	.17	.13*	.22	.56	.27	.28	.60	.51
Hostility	.41	.30	.31	.10*	.21	.05^{ns}	.39	.15**	.20	.42	.36
Phobic Anxiety	.39	.14**	.39	.13*	.15**	.13*	.39	.15**	.21	.42	.37
Paranoia	.48	.28	.49	.18	.21	.21	.54	.30	.33	.55	.51
Psychoticism	.49	.19	.48	.09^{ns}	.16**	.19	.49	.17	.22	.51	.43
Global Severity Index	.54	.24	.55	.16	.20	.21	.57	.25	.29	.59	.51
Obsessive–Compulsive Inventory[d]											
Frequency	.43	.24	.45	.39	.08^{ns}	.34	.52	.42	.47	.47	.54
Distress	.50	.28	.49	.40	.03^{ns}	.33	.60	.44	.48	.51	.57
Fear of Negative Evaluation[a]	.63	.26	.83	.16	.20	.31	.64	.33	.34	.73	.62
Social Desirability: MCSDS[c]	–.15**	–.17	–.09*	–.04^{ns}	–.14**	–.09*	–.18	–.16	–.12**	–.18	–.18

Notes. For all correlations, $p < .001$ (except as noted). CM = Concern Over Mistakes; HS = High Standards for Others; NA = Need for Approval; OR = Organization; PP = Perceived Parental Pressure; PL = Planfulness; RU = Rumination; SE = Striving for Excellence; CP = Conscientious Perfectionism; SEP = Self-Evaluative Perfectionism; PI–C = Perfectionism Inventory Composite score; MPS–F = Frost's Multidimensional Perfectionism Scale; MPS–HF = Hewitt and Flett's Multidimensional Perfectionism Scale; BSI = Brief Symptom Index; MCSDS = Marlowe–Crowne Social Desirability Scale.

[a]$n = 613$. [b]$n = 355$. [c]$n = 368$. [d]$n = 207$.

*$p < .05$, one-tailed. **$p < .01$, one-tailed. $^{ns}p > .05$, all one-tailed.

Example: Perfectionism

- Examination of the pattern of correlations revealed convergent and discriminate evidence for the PI:
  - Concern Over Mistakes scale of the PI strongly associated with a Concern Over Mistakes Scale from a different measure of perfectionism
  - Striving For Excellence scale of the PI strongly associated with Personal Standards scale and a Self-Oriented Perfectionism scale
  - Rumination, Concern Over Mistakes, and Need For Approval scales of PI strongly associated with fear of negative evaluation, frequency, and severity of symptoms of obsessive-compulsive disorder

Example: Perfectionism

- Examination of the pattern of correlations revealed convergent and discriminate evidence for the PI:
  - Concern Over Mistakes scale of the PI strongly associated with a Concern Over Mistakes Scale from a different measure of perfectionism
  - Striving For Excellence scale of the PI strongly associated with Personal Standards scale and a Self-Oriented Perfectionism scale
  - Rumination, Concern Over Mistakes, and Need For Approval scales of PI strongly associated with fear of negative evaluation, frequency, and severity of symptoms of obsessive-compulsive disorder

Example: Perfectionism

Correlations Between Perfectionism Inventory Scales and Related Measures

Scale	*CM*	*HS*	*NA*	*OR*	*PP*	*PL*	*RU*	*SE*	*CP*	*SEP*	*PL–C*
Perfectionism: MPS–F[a]											
Concern Over Mistakes	.82	.43	.58	.18	.38	.30	.70	.52	.47	.78	.72
Doubts About Actions	.63	.37	.60	.24	.20	.38	.70	.43	.47	.67	.65
Parental Criticism	.41	.25	.20	–.03^{ns}	.60	.02^{ns}	.32	.17	.14	.49	.36
Parental Expectations	.31	.27	.18	.07^{ns}	.85	.06^{ns}	.29	.32	.23	.53	.43
Personal Standards	.47	.50	.36	.45	.39	.44	.52	.72	.70	.55	.71
Organization	.12	.36	.18	.89	.11**	.49	.31	.51	.76	.23	.55
Perfectionism: MPS–HF[b]											
Self-Oriented	.47	.42	.34	.47	.42	.45	.55	.79	.71	.57	.73
Other-Oriented	.33	.62	.14**	.29	.30	.26	.37	.42	.53	.36	.51
Socially-Prescribed	.65	.35	.49	.16**	.58	.21	.61	.42	.38	.74	.65
Symptoms: BSI[c]											
Somatic Complaints	.35	.14*	.31	.13*	.11*	.13*	.34	.17	.19	.35	.31
Depression	.46	.16**	.46	.03^{ns}	.15**	.18	.46	.13*	.17	.49	.39
Obsessive–Compulsive	.40	.14**	.46	.08^{ns}	.10*	.19	.46	.18	.19	.45	.37
Anxiety	.42	.28	.42	.22	.25	.25	.49	.29	.35	.50	.49
Interpersonal Sensitivity	.52	.18	.68	.17	.13*	.22	.56	.27	.28	.60	.51
Hostility	.41	.30	.31	.10*	.21	.05^{ns}	.39	.15**	.20	.42	.36
Phobic Anxiety	.39	.14**	.39	.13*	.15**	.13*	.39	.15**	.21	.42	.37
Paranoia	.48	.28	.49	.18	.21	.21	.54	.30	.33	.55	.51
Psychoticism	.49	.19	.48	.09^{ns}	.16**	.19	.49	.17	.22	.51	.43
Global Severity Index	.54	.24	.55	.16	.20	.21	.57	.25	.29	.59	.51
Obsessive–Compulsive Inventory[d]											
Frequency	.43	.24	.45	.39	.08^{ns}	.34	.52	.42	.47	.47	.54
Distress	.50	.28	.49	.40	.03^{ns}	.33	.60	.44	.48	.51	.57
Fear of Negative Evaluation[a]	.63	.26	.83	.16	.20	.31	.64	.33	.34	.73	.62
Social Desirability: MCSDS[c]	–.15**	–.17	–.09*	–.04^{ns}	–.14**	–.09*	–.18	–.16	–.12**	–.18	–.18

Notes. For all correlations, $p < .001$ (except as noted). CM = Concern Over Mistakes; HS = High Standards for Others; NA = Need for Approval; OR = Organization; PP = Perceived Parental Pressure; PL = Planfulness; RU = Rumination; SE = Striving for Excellence; CP = Conscientious Perfectionism; SEP = Self-Evaluative Perfectionism; PI–C = Perfectionism Inventory Composite score; MPS–F = Frost's Multidimensional Perfectionism Scale; MPS–HF = Hewitt and Flett's Multidimensional Perfectionism Scale; BSI = Brief Symptom Index; MCSDS = Marlowe–Crowne Social Desirability Scale.

[a]$n = 613$. [b]$n = 355$. [c]$n = 368$. [d]$n = 207$.

*$p < .05$, one-tailed. **$p < .01$, one-tailed. $^{ns}p > .05$, all one-tailed.

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Multitrait–Multimethod Matrices

- Cronbach and Meehl described the concept of construct validity
- However, they did not fully describe a method to test construct validity
- Campbell and Fiske developed the logic of the multitrait-multimethod matrix as a statistical extension of Cronbach and Meehl's work
- The main problem the MTMM tries to overcome is the fact that a correlation between two scores may conflate two sources of variance:
  - **trait variance** (the good stuff)
  - **method variance** (the bad stuff)

Multitrait–Multimethod Matrices

- Researchers who use the MTMM approach to validate the interpretations of test scores must administer their inventory using different methods
  - at least three, in practice
- Essentially, the MTMM approach is based on the notion that we "hope" to see larger correlations between scores based on the same traits (irrespective of method of measurement), in comparison to correlations between scores based simply on the same method

Multitrait–Multimethod Matrices

- That is, large correlations between different traits using the same measurement method are not interesting theoretically
- It suggests that the correlations are simply due a response style (i.e., method variance)
- We want a lot shared trait variance, particularly identical traits measured using different methods

Multitrait–Multimethod Matrices

- For example, consider a measure of *social skill* that we may hope to validate
- We could administer the questionnaire in addition to other measures such as a measure of *impulsivity*, *conscientiousness*, and *emotional stability*
- Theoretically, we would expect some "small-ish" correlations between social skill and these other three measures
- This would be the **multitrait** component of the study

Multitrait–Multimethod Matrices

- In addition to collecting data via the *self-report method*, we could also use *acquaintance reports*
- That is, each person who completed the self-report questionnaires would have someone that knows them well to complete the same questionnaires phrased in the third person
- Also, the four traits (social skill, impulsivity, conscientiousness, emotional stability) could also be measured using an *interview based technique*
- Thus, there would be three methods of measurement: self-report, rater-report, and interview—the **multimethod** component of the study

Multitrait–Multimethod Matrices: Four Types of Associations

Association Between the Two Constructs		*Method Used to Measure the Two Constructs*	
		Different Methods (e.g., Self-Report for One Construct and Acquaintance Report for the Other)	***Same Method (e.g., Self-Report Used for Both Constructs)***
Different constructs (not associated)	Label	Heterotrait-heteromethod correlations	Heterotrait-monomethod correlations
	Sources of variance	Nonshared trait variance and nonshared method variance	Nonshared trait variance and shared method variance
	Example	Self-report measure of social skill correlated with acquaintance report measure of emotional stability	Self-report measure of social skill correlated with self-report measure of emotional stability
	Expected correlation	Weakest	Moderate?
Same (or similar) constructs (associated)	Label	Monotrait-heteromethod correlations	Monotrait-monomethod correlations
	Sources of variance	Shared trait variance and nonshared method variance	Shared trait variance and shared method variance
	Example	Self-report measure of social skill correlated with acquaintance report measure of social skill	Self-report measure of social skill correlated with self-report measure of social skill (i.e., reliability)
	Expected correlation	Moderate?	Strongest

Multitrait–Multimethod Matrices

Methods	Traits	Self-Report: Social Skill	Self-Report: Impulsivity	Self-Report: Conscientiousness	Self-Report: Emotional Stability	Acquaintance Report: Social Skill	Acquaintance Report: Impulsivity	Acquaintance Report: Conscientiousness	Acquaintance Report: Emotional Stability	Acquaintance Report: Social Skill	Acquaintance Report: Impulsivity	Acquaintance Report: Conscientiousness	Acquaintance Report: Emotional Stability
Self-report	Social Skill	(.85)											
	Impulsivity	.14	(.81)										
	Conscientiousness	.20	.22	(.75)									
	Emotional Stability	.35	.24	.19	(.82)								
Acquaintance	Social Skill	**.40**	.14	.10	.22	(.76)							
	Impulsivity	.13	**.32**	.13	.19	.18	(.80)						
	Conscientiousness	.09	.17	**.36**	.14	.14	.26	(.68)					
	Emotional Stability	.20	.23	.11	**.41**	.30	.28	.18	(.78)				
Interviewer report	Social Skill	**.34**	.11	.19	.20	**.23**	.01	.11	.19	(.81)			
	Impulsivity	.03	**.25**	.12	.19	.06	**.24**	.10	.14	.22	(.77)		
	Conscientiousness	.09	.09	**.30**	.14	.09	.08	**.20**	.06	.24	.30	(.86)	
	Emotional Stability	.14	.16	.08	**.33**	.13	.12	.06	**.19**	.44	.38	.29	(.78)

Heterotrait-Heteromethod

Heterotrait-Monomethod

Monotrait-Heteromethod (Validity)

Monotrait-Monomethod (Reliability)

Multitrait–Multimethod Matrices

The most stringent test in a MTMM analysis is to determine whether the monotrait-heteromethod correlations are "meaningfully" larger than the heterotrait-monomethod correlations

With respect to social skill, the mean "monotrait-heteromethod" correlation is equal to **.32** ([.40 + .34 + .23] / 3)

Methods	*Traits*	*Self-Report*				*Acquaintance Report*				*Acquaintance Report*			
		Social Skill	*Impulsivity*	*Conscien-tiousness*	*Emotional Stability*	*Social Skill*	*Impulsivity*	*Conscien-tiousness*	*Emotional Stability*	*Social Skill*	*Impulsivity*	*Conscien-tiousness*	*Emotio Stabilit*
Self-report	Social Skill	(.85)											
	Impulsivity	.14	(.81)										
	Conscientiousness	.20	.22	(.75)									
	Emotional Stability	.35	.24	.19	(.82)								
Acquaintance	Social Skill	**.40**	.14	.10	.22	(.76)							
	Impulsivity	.13	**.32**	.13	.19	.18	(.80)						
	Conscientiousness	.09	.17	**.36**	.14	.14	.26	(.68)					
	Emotional Stability	.20	.23	.11	**.41**	.30	.28	.18	(.78)				
Interviewer report	Social Skill	**.34**	.11	.19	.20	**.23**	.01	.11	.19	(.81)			
	Impulsivity	.03	**.25**	.12	.19	.06	**.24**	.10	.14	.22	(.77)		
	Conscientiousness	.09	.09	**.30**	.14	.09	.08	**.20**	.06	.24	.30	(.86)	
	Emotional Stability	.14	.16	.08	**.33**	.13	.12	.06	**.19**	.44	.38	.29	(.78)

Heterotrait-Heteromethod
Heterotrait-Monomethod
Monotrait-Heteromethod (Validity)
Monotrait-Monomethod (Reliability)

Multitrait–Multimethod Matrices

The most stringent test in a MTMM analysis is to determine whether the monotrait-heteromethod correlations are "meaningfully" larger than the heterotrait-monomethod correlations

By contrast, the mean self-report "heterotrait-monomethod" correlation is equal to **.22** ([.14 + .20 + .35 + .22 + .24 + .19] / 6)

Methods	Traits	Self-Report: Social Skill	Self-Report: Impulsivity	Self-Report: Conscientiousness	Self-Report: Emotional Stability	Acquaintance Report: Social Skill	Acquaintance Report: Impulsivity	Acquaintance Report: Conscientiousness	Acquaintance Report: Emotional Stability	Acquaintance Report: Social Skill	Acquaintance Report: Impulsivity	Acquaintance Report: Conscientiousness	Acquaintance Report: Emotional Stability
Self-report	Social Skill	(.85)											
	Impulsivity	.14	(.81)										
	Conscientiousness	.20	.22	(.75)									
	Emotional Stability	.35	.24	.19	(.82)								
Acquaintance	Social Skill	**.40**	.14	.10	.22	(.76)							
	Impulsivity	.13	**.32**	.13	.19	.18	(.80)						
	Conscientiousness	.09	.17	**.36**	.14	.14	.26	(.68)					
	Emotional Stability	.20	.23	.11	**.41**	.30	.28	.18	(.78)				
Interviewer report	Social Skill	**.34**	.11	.19	.20	**.23**	.01	.11	.19	(.81)			
	Impulsivity	.03	**.25**	.12	.19	.06	**.24**	.10	.14	.22	(.77)		
	Conscientiousness	.09	.09	**.30**	.14	.09	.08	**.20**	.06	.24	.30	(.86)	
	Emotional Stability	.14	.16	.08	**.33**	.13	.12	.06	**.19**	.44	.38	.29	(.78)

Heterotrait-Heteromethod
Heterotrait-Monomethod
Monotrait-Heteromethod (Validity)
Monotrait-Monomethod (Reliability)

Multitrait–Multimethod Matrices

- Is the difference between **.32** and **.22** larger enough to merit a favourable evaluation of the social skill measure?
- That is one of the limitations associated with this approach
- There are no clear guidelines to evaluate the differences in the mean correlations
- At the very least, the monotrait-heteromethod correlations need to be larger than the heterotrait-monomethod correlations
- The lack of guidelines is probably one of the main reasons why MTMM studies are rare
- The other reason would be that they are labour intensive to conduct

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Four Methods For Evaluating Convergent and Discriminant Validity

- There are four methods used to evaluate the degree to which measures show convergent and discriminant associations
  1. Focused associations
  2. Sets of correlations
  3. Multitrait–Multimethod Matrices
  4. Quantifying Construct Validity

Quantifying Construct Validity

- The methods described thus far are rather imprecise and subjective approaches to evaluating a pattern of convergent and discriminant correlations
- Westen and Rosenthal (2003) outlined a more precise and objective quantitative procedure called *quantifying construct validity* (QCV)
- Essentially, this procedure requires researchers to predict the magnitude of the correlation between their measure of interest and their selected criteria
- Then, the correlations between their measure of interest and these selected criteria are estimated
- Finally, the correlation between the predicted and estimated correlations is estimated

Example: Social Motivation

- Furr et al. (2004) examined the construct validity associated with a new measure of social motivation—a person's general desire to make positive impressions on other people
- The researchers had a group of five experts predict what the correlation would be between the measure of social motivation and 12 other self-report measures of personality like attributes (e.g., self-efficacy, agreeableness, need to belong)
- They then took the average of the estimates
- Next, they got people to respond to the questionnaires and estimated the empirical correlations

Example: Social Motivation

Criteria Scales	Predicted Correlations	Actual Correlations	z-Transformed Correlations
Dependence	.58	.46	.50
Machiavellianism	.24	.13	.13
Distrust	–.04	–.24	–.24
Resourcefulness	.06	–.03	–.03
Self-efficacy	–.04	.12	.12
Extraversion	.18	.03	.03
Agreeableness	.36	.39	.41
Complexity	.08	.06	.06
Public self-consciousness	.64	.51	.56
Self-monitoring	.56	.08	.08
Anxiety	.36	.24	.24
Need to belong	.56	.66	.79

Example: Social Motivation

- The Pearson correlation between the "Predicted Correlations" and the "Actual Correlations" is equal to .79, a large positive correlation
- However, at this point in time, there are no clear guidelines regarding how large the correlation should be to be interpreted as providing evidence of adequate validity
- All we can say is that higher correlations offer greater evidence of validity
- The correlation of .79 would seem to indicate a high degree of convergent and discriminant validity

Quantifying Construct Validity

- The QCV approach has several advantages:
  1. It forces researchers to consider carefully the expected pattern of convergent and discriminant associations that would make theoretical sense
  2. It forces researchers to make explicit quantitative predictions about the pattern of associations
  3. It provides a single value reflecting the overall "goodness-of-fit" between the predicted and actual associations

Quantifying Construct Validity

- The QCV approach has several advantages:
  1. It forces researchers to consider carefully the expected pattern of convergent and discriminant associations that would make theoretical sense
  2. It forces researchers to make explicit quantitative predictions about the pattern of associations
  3. It provides a single value reflecting the overall "goodness-of-fit" between the predicted and actual associations

Quantifying Construct Validity

- The QCV approach has several advantages:
  1. It forces researchers to consider carefully the expected pattern of convergent and discriminant associations that would make theoretical sense
  2. It forces researchers to make explicit quantitative predictions about the pattern of associations
  3. It provides a single value reflecting the overall "goodness-of-fit" between the predicted and actual associations

Quantifying Construct Validity

- The QCV approach has several advantages:
  1. It forces researchers to consider carefully the expected pattern of convergent and discriminant associations that would make theoretical sense
  2. It forces researchers to make explicit quantitative predictions about the pattern of associations
  3. It provides a single value reflecting the overall "goodness-of-fit" between the predicted and actual associations

Quantifying Construct Validity

- However, the approach is not without its limitations
- A low correlation between the predicted and actual associations does not necessarily reflect poor validity—the predicted associations may simply be a poor reflection of a construct's nomological network
- Conversely, a high correlation between predicted and actual associations does not necessarily reflect good validity—it is possible to obtain a relatively large correlation when the predicted and actual associations do not closely match
- Some care is therefore required in interpreting the results of a QCV analysis

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

True Associations Between Constructs

- Recall from our week 4 lecture that one factor affecting the correlation between measures of two constructs is the "true" association between those constructs
- If two constructs are strongly associated with each other, then measures of those constructs will likely be highly correlated with each other
- Conversely, if two constructs are unrelated to each other, then measures of those constructs will probably be weakly correlated with each other
- We tend to interpret correlations between measured variables as approximations of the true associations between the constructs we are interested in

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

Measurement Error and Reliability

- Recall from our week 4 lecture that the correlation between measures of two constructs is not only affected by the true association between those constructs
- It is also affected by measurement error
- Hence, the correlation between two measures is a function of the true correlation and the reliabilities of the two tests:

$$r_{x_o y_o} = r_{x_t y_t} \sqrt{R_{xx} R_{yy}} \tag{21}$$

- That is, measurement error reduces—or "attenuates"—the correlation between measures
- Measurement error therefore affects validity coefficients, just like any other correlation

Measurement Error and Reliability

- Researchers are generally satisfied if a test's reliability is above .70 or .80
- If a test's or a criterion's reliability is much lower than .70, then we should have concerns about its effect on a validity coefficient
- In this case there are two options:
  1. disregard—or reduce the weight given to—a validity coefficient based on poor reliability
  2. adjust the validity coefficient to account for measurement error

Measurement Error and Reliability

- As discussed in our week 4 lecture—and week 5 lab—correlation coefficients can be "disattentuated" for imperfect reliability using the *correction for attenuation* formula:

$$r_{x_t y_t} = \frac{r_{x_o y_o}}{\sqrt{R_{xx} R_{yy}}} \tag{22}$$

- This adjusts a validity correlation by assuming that both variables were measured without any measurement error

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

Factors Affecting Validity

- There are at least three factors that affect the size of validity coefficients:
  1. True associations between constructs
  2. Measurement error and reliability
  3. Restricted range

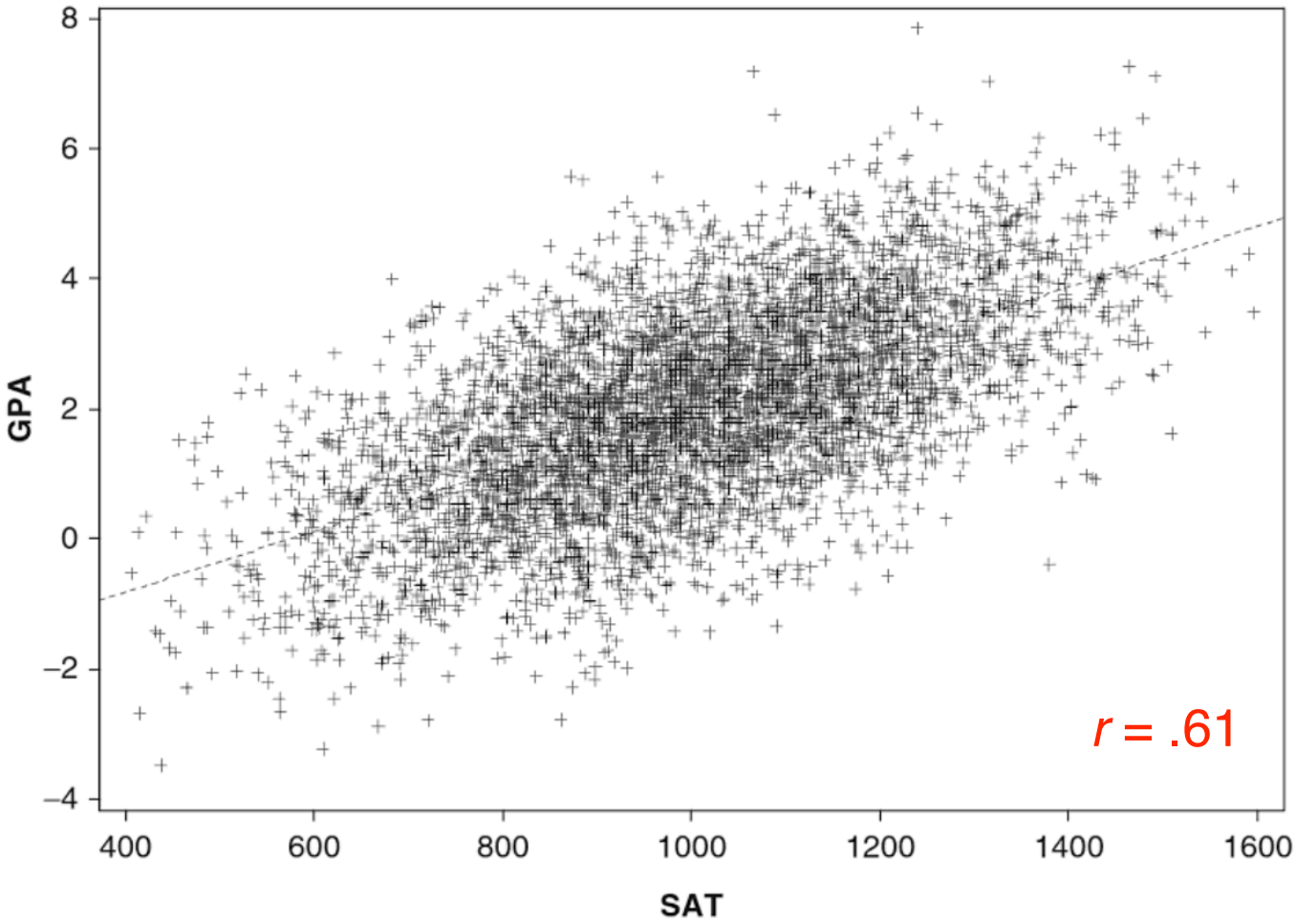
Restricted Range

- The amount of variability in one or both distributions of scores can affect the correlation between the two sets of scores
- Specifically, a correlation between two variables can be reduced if the range of scores in one or both variables is artificially limited or restricted

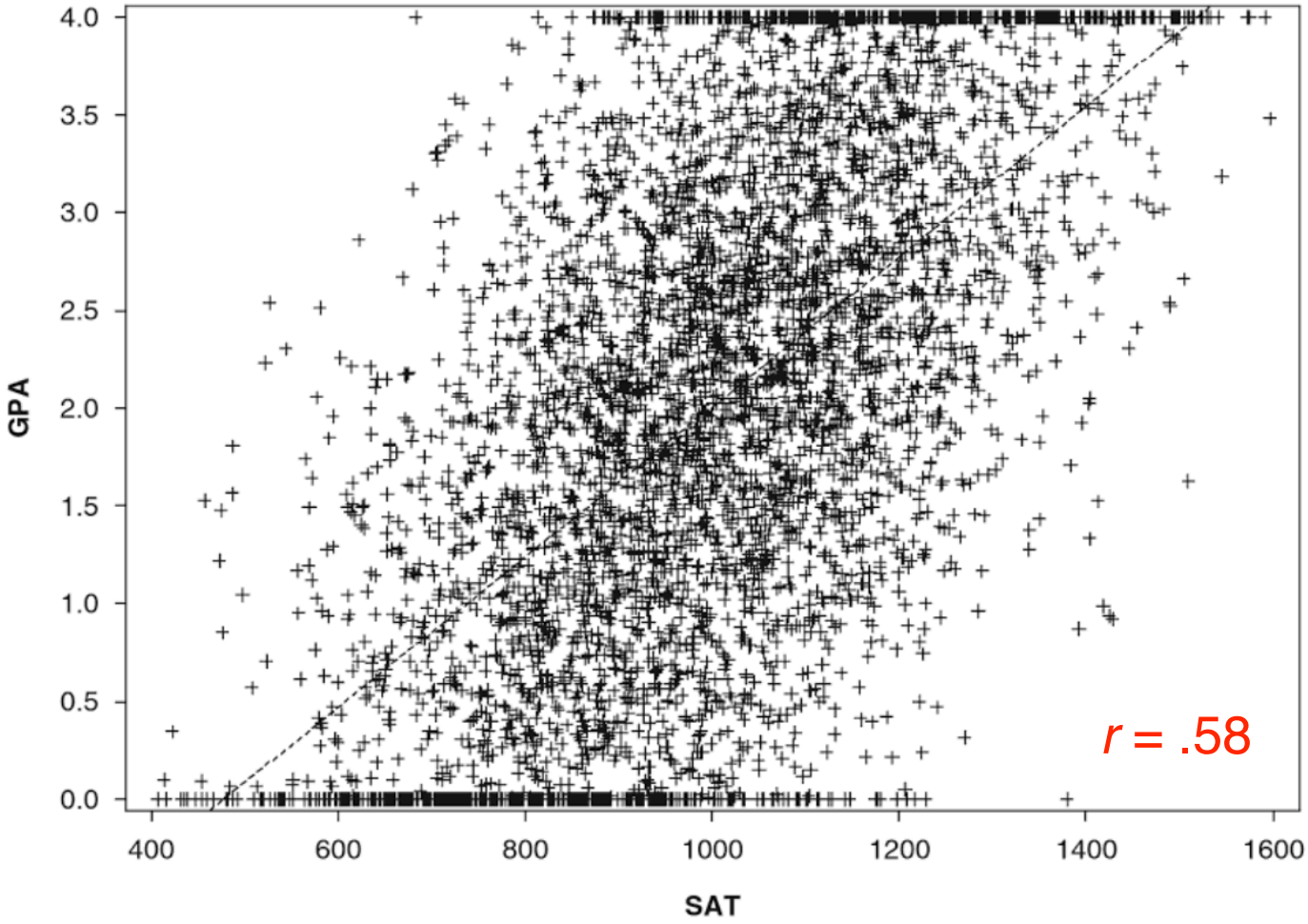
Restricted Range

- The textbook gives the example of the correlation between SAT and GPA scores to illustrate restricted range
- Imagine a scenario where two students achieve a Grade Point Average of 4.0 (remember GPA scores vary between 0 and 4
- As an A grade indicates an 80 or higher, it is possible for someone with an average of 81 to have a GPA of 4.0 and another person with an average of 91 to have a GPA of 4.0
- If we were to use GPA as the dependent variable, it would constrain the amount of variance that could be used in a correlation

Restricted Range: Scatterplot of SAT and "Unrestricted" GPA



Restricted Range: Scatterplot of SAT and "Restricted" GPA



Restricted Range

- Range restriction can be difficult to diagnose
- If the range of obtained scores is dramatically different from the range of possible scores, then there might be a range restriction issue
- If the range of obtained scores falls to one "side" of the distribution of possible scores, then there might be serious concerns about range restrictions
- In the example, the impact is relatively small (from r = .61 to r = .58), but you should know that the effects can be much more pronounced
- There are no easy tricks for detecting range restriction, but it is a problem that you need to be aware of

References

Furr, M. R., & Bacharach, V. R. (2014; Chapter 9). Psychometrics: An Introduction (second edition). Sage.