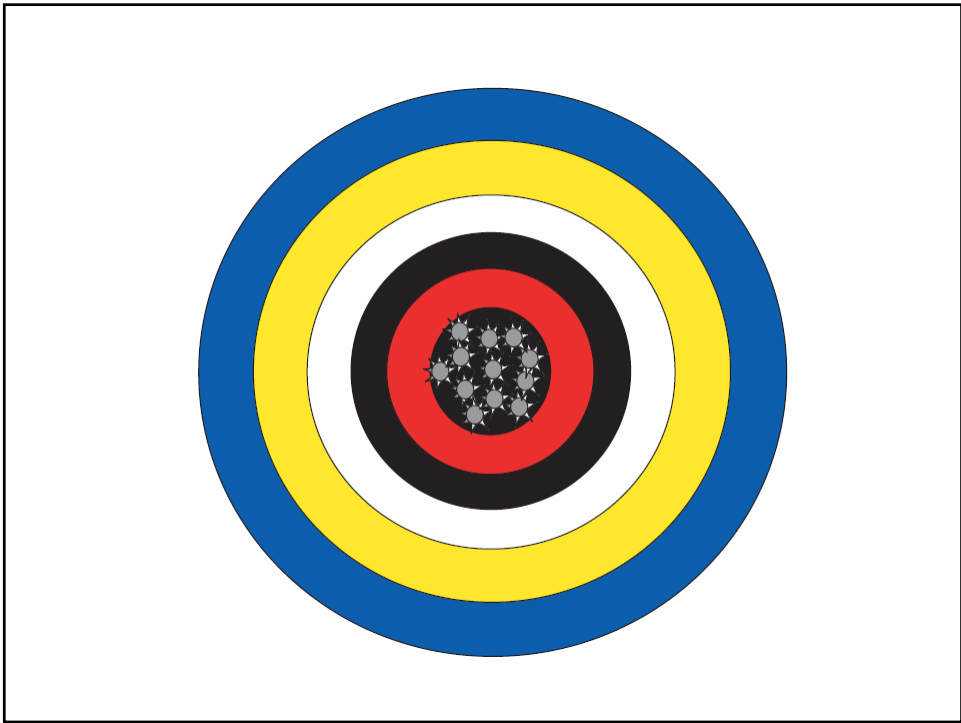
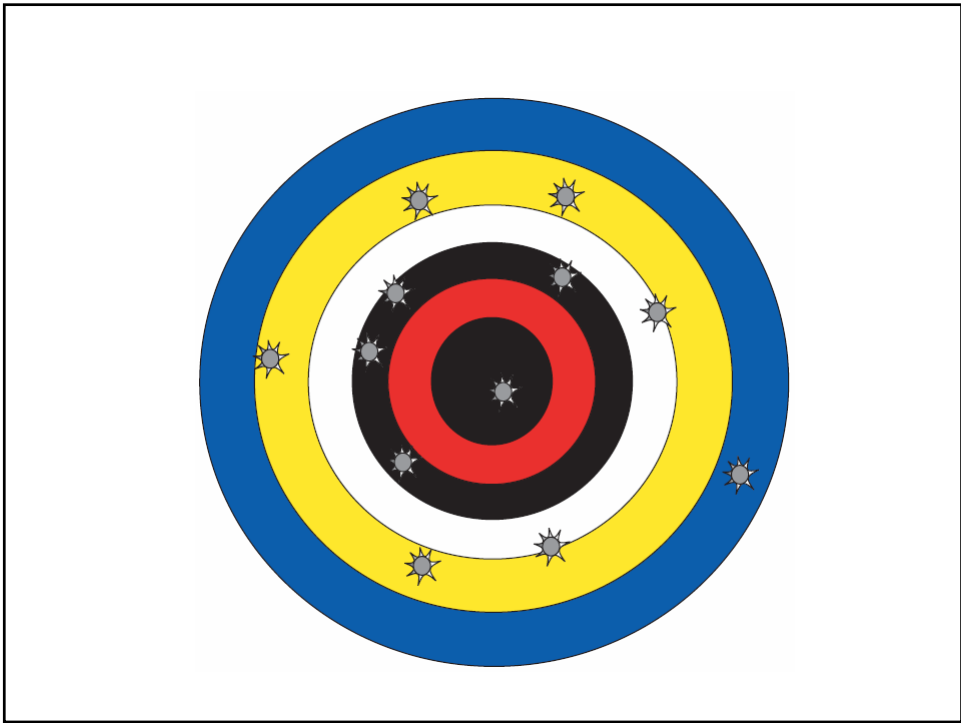


## II. Measurement Validity

### A. Validity and Reliability Hitting the Target







- Beyond Consistency we should be concerned about whether our measures identify what they are supposed to.

B. The Classic Triumvirate

**Content Validity** – Experts' Eyeballs

**Criterion Validity** – From Measures to Outcomes

**Content Validity** – What Would Theory Do?

- Together, these inform us of Validity.
- Validity is a Unified Construct

## 1. Content Validity (Face Validity):

- Do the items appear to tap what they are supposed to.
- Assessed by SMEs (Subject Matter Experts)
  - Relevance: only items relevant to a construct are included.
  - Representativeness: all relevant dimensions are included.

## 2. Criterion Validity / Empirical Validity

- measure predicts results on relevant, validated instrument (criterion)
  - specific behavior
  - different operationalization of construct
- **Concurrent** – criterion and predictor assessed at same time
- **Predictive** – assess criterion at a later time

### 3. Construct Validity

Does measure tap the construct of interest

#### a. 3 steps

a1. **Hypothesize** – Theory outlines relation with other constructs

#### a2. **Test**

- Correlation with other valid constructs
- Experimental/Quasi-Experimental – test role of measure as moderator of relevant manipulations

#### Examples



Correlation – GSR x SHeMALE (Scale of Heavy Meatal Attitudes: Long Edition)  
(Nugent et al. 1979)

Experiment/Quasi-Experiment: The Barbie/Ken Dressing Activity Game.  
(formerly the B/K Genderedness Assessment for Youths: B. George & R. Paule, 1988)

P's given either B or K doll along with doll sized male and female clothing.

Group	Low GSR	High GSR
Barbie	Dress Doll in Female Clothing	Ate Doll and Tried to wear female doll clothing
Ken	Dress Doll in Male Clothing	Ate Clothing & Cut off head, hands, and left foot of doll in ritual sacrifice to Odin.

a3. **Infer** construct validity from results

- Reject Null = evidence supporting construct validity
- Fail to Reject Null =
  1. could be lack of construct validity
  2. method could be flawed (Power problems, threats to internal validity).
  3. Theory could be flawed
  4. Measures of other constructs may not be valid

b. Known Groups Method

- Predict direction of differences between groups with known attributes

e.g. GSR scores for

Kiss Fans vs. Dave Mathews Fans



Base Players Vs. Non-Base Players

Kiss Fans vs. Britney Spears Fans (no difference found)

e. Structural Validity

- Use confirmatory factor analysis to determine whether items are consistent with the predicted dimensions.

f. Convergent Validity

- Demonstrate appropriate associations with constructs related by theory

e.g. GSR x SHeMale :  $r = .50$

GSR x BGR (Bill Gates Resemblance):  $r = -.50$

- Indicate High degree of association but not redundant (e.g. .70 or above)



Friends don't let Friends Dress like Bill Gates – Only you can prevent BGR

### g. Discriminant/Divergent Validity

- Measure is uncorrelated (weakly correlated) with theoretically distinct/unrelated constructs.

- The Null hypothesis problem – You need adequate power (.95) to ensure that lack of association is not simply Type II error.

Example – GSR x Neuroticism :  $r = .18$   
GSR x Extroversion :  $r = .02$

(Big Five: Costa & McCrea)

### Putting it all Together

- The MTMX –
  - Multi-Trait, Multi-Method Matrix (Campbell & Fiske, 1959)
    - Evaluate Criterion Related Validity, Convergent Validity, and Divergent/Discriminant Validity
  - Multi-Trait
    - To assess new measure use 2+ theoretically distinct constructs
      - convergent & divergent
  - Multi-Method
    - Measure each construct multiple ways
      - Use as many methods as you have constructs (results in a square matrix)
      - Alternate methods can be references for criterion behavior
      - Helps Identify Common/Shared Method Variance

- Example
- Measure of interest GSR-self report (GSRsr)
- Alternate Methods – peer report (GSRpr), diary behavior report (GSRdr).
- Alternate Traits
  - Convergent – SHeMALE-self report (SHeMALEsr), peer report (pr), and diary behavior report (dr)
  - Divergent – Neuroticism – self report (Nsr), peer report (Npr), and diary behavior report (Ndr).

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.59	0.55	0.66	1.00					
SHeMale pr	0.65	0.60	0.65	0.86	1.00				
SHeMale dr	0.53	0.69	0.61	0.85	0.84	1.00			
N sr	0.06	0.01	0.03	-0.06	0.11	0.08	1.00		
N pr	0.10	0.02	-0.03	0.05	0.12	-0.03	0.85	1.00	
N dr	-0.11	0.12	0.10	0.15	-0.09	0.07	0.86	0.90	1.00

-Criterion Validity of GSR-sr

-GSRsr highly correlated with GSR assessed by other methods

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.59	0.55	0.66	1.00					
SHeMale pr	0.65	0.60	0.65	0.86	1.00				
SHeMale dr	0.53	0.69	0.61	0.85	0.84	1.00			
N sr	0.06	0.01	0.03	-0.06	0.11	0.08	1.00		
N pr	0.10	0.02	-0.03	0.05	0.12	-0.03	0.85	1.00	
N dr	-0.11	0.12	0.10	0.15	-0.09	0.07	0.86	0.90	1.00

-Convergent Validity of construct

-GSR strongly correlated SHeMALE construct – across all Methods of assessment

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.59	0.55	0.66	1.00					
SHeMale pr	0.65	0.60	0.65	0.86	1.00				
SHeMale dr	0.53	0.69	0.61	0.85	0.84	1.00			
N sr	0.06	0.01	0.03	-0.06	0.11	0.08	1.00		
N pr	0.10	0.02	-0.03	0.05	0.12	-0.03	0.85	1.00	
N dr	-0.11	0.12	0.10	0.15	-0.09	0.07	0.86	0.90	1.00

- Divergent Validity of Construct
- GSR not meaningfully associated with Neuroticism – across all methods of assessment

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.59	0.55	0.66	1.00					
SHeMale pr	0.65	0.60	0.65	0.86	1.00				
SHeMale dr	0.53	0.69	0.61	0.85	0.84	1.00			
N sr	0.06	0.01	0.03	-0.06	0.11	0.08	1.00		
N pr	0.10	0.02	-0.03	0.05	0.12	-0.03	0.85	1.00	
N dr	-0.11	0.12	0.10	0.15	-0.09	0.07	0.86	0.90	1.00

### -Common Method Variance

-When correlations between constructs is attributable to the method used and not the constructs themselves

-In this example correlations between different constructs are low, except when the constructs are assessed using the same method (highlighted in pink)

-In studies where all self-report measures are used, it is difficult to distinguish between common method variance and true relationships between constructs.

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.45	0.06	0.16	1.00					
SHeMale pr	0.20	0.40	0.06	0.86	1.00				
SHeMale dr	0.15	0.15	0.48	0.85	0.84	1.00			
N sr	0.55	0.01	0.03	0.45	0.11	0.08	1.00		
N pr	0.10	0.50	-0.03	0.05	0.43	-0.03	0.85	1.00	
N dr	-0.11	0.12	0.44	0.15	-0.09	0.55	0.86	0.90	1.00

-Lack of divergence

-The Strong Associations between the GSR measures and the Neuroticism measures suggests that GSR may not be conceptually distinct from Neuroticism as theory would suggest.

-It may be that Heavy Metal attitudes and Neuroticism represent orthogonal factors that make up GSR.

-We may want to rethink our GSR model and conduct more studies that test hypotheses derived from this revised model.

	GSR sr	GSR pr	GSR dr	SHeMale sr	SHeMale pr	SHeMale dr	N sr	N pr	N dr
GSR sr	1.00								
GSR pr	0.80	1.00							
GSR dr	0.83	0.88	1.00						
SHeMale sr	0.49	0.45	0.56	1.00					
SHeMale pr	0.55	0.50	0.55	0.86	1.00				
SHeMale dr	0.43	0.49	0.51	0.85	0.84	1.00			
N sr	0.53	0.33	0.32	-0.06	0.11	0.08	1.00		
N pr	0.52	0.24	0.46	0.05	0.12	-0.03	0.85	1.00	
N dr	0.43	0.37	0.58	0.15	-0.09	0.07	0.86	0.90	1.00