# Validity

- Reliability: degree of consistency in measurement
- Validity: Degree of accuracy in measurement

**Construct validity**

- does a test measure the construct that it purports to measure?
  - content validity
  - convergent validity
  - divergent validity

**Criterion-related validity**

- Does the measure predict any critieria of interest
  - meausre predict useful criteria
- incremetnal validity
- predictive/concurrent validity
- validity generalization

## Construct validity

- checking if the acutal content of the measure
- a measure has strong content validity if the items adequately reflect the construct of interest

**content validity**

- content valid – contain adequate sampling of the full range of behaviours that represent the construct
  - e.g. poor content validity: material not covered in class/not adequately represented on the exam
- content validity is assessed via subjective judgements
  - judgements are made by experts
  - judgements are averaged across serveral non-experts
- The content of the measure is evaluated as to the degree to which the items reflect the construct.

**Pattern of correlations**

- expect strong correlations with: (convergent validity)
  - different measuers of the same construct
  - measures of conceptually similar constructs
- expect weak correlation with: (divergent validity)
  - measures of conceptually dissimilar constructs
- e.g. measure of industriousness
  - measure of orderliness (strongly correlated)
  - measure of enthusiasm (weakly correlated)

**Correlation matrix**

- correlation matrices are symmetrical
  - typically one of the "triangles" is empty in a correlation matrix
- reliabilities are displayd on the "diagnoal"
- contain descriptive statisitics
  - e.g. means and s.d.
- contain correlations between all variables include in a study
  - the information we need to access the convergent and divergent validity
  - positive: strongly correlated $\rightarrow$ convergent validity
  - negative: weakly correlated $\rightarrow$ divergent validity
- Taken together, the pattern of correlations provides evidence for the construct validity of our measure

## Crieterion-related validity

- criterion-related validity is the degree to which our measure predicts a criterion of interest
  - intelligence predict task performance
  - work interests predict turnover
  - integrit predict CWBs

- **Validity coefficient**: a correlation between a measure and a criterion
  - validity coefficient tells us how good of a predictor a measure is for a given criterion
  - the strong positive correlation indicates that the higher an applicant scores on the industriousness measure, the more successful they are likely to perform in their university courses

### Incremental validity

- The degree to which a predictor accounts for unique variance in the dependent variable above and beyond other predictors
- When we add a predictor, we want to make sure it will account for **incremental variance** in task performance
  - incremental variance is the unique variance accounted for by each predictor
- achieve high levels of incremental validity
  - Add predictors that provide unique information
  - each individual predictors is not correlated with each other to achieve the highest levels of incremental validity
  - multiple predictors with each predictor has strong correlation bewteen itself and criteria while weak correlations among the predictors

## Combine predictors

- optimal weighting
- unit weighting

### Optimal weighting

- multiple regression can be used to find the optimal weights for combining predictors
  - MR finds the weights that maximizes the $R_{xy}$ for that sample
  - Predictor scores are combined into a weighted linear compositepredictor.
- downside of using optimal weights is that the weights are only optimal for that specific sample.
- $R_{xy}$ from the original sample will always be larger than the $R_{xy}$ we observe when we actually use the weights during hiring
  - shrinkage: $R_{xy}$ "shrinks" in the applicant sample

**Unit weighting**

- multiple predictors is far less susceptible to shrinkage
- Procedure:
  - transform to $z - scores$
  - unit-weighted composite scores for $n$ participants

**When to use optimal-unit-weighting**

- Depends on the sample size
  - If sample size is very large(hundreds or thousands), optimal-weighting precferred
  - If sample size is small, unit-weighting preferred

# Where does criterion-related validity evidence come from

- collect both predictor and criterion data to see if predictors will be useful for hiring in the future
- Need a sample of participants to provide both predictor and criterion data

## Concurrent validity studies

- Predictors and criteria collected at the same time(concurrently)
- Sample = **current employees**
  - Administer predictor measures
  - Correlate with available criterion data
- Advantage:
  - data are available immediately
- disadvantage:
  - not the sample of interest
    - use predictors with applicants during selection
    - applicant may be different from current employees
  - restricted range
    - employees were the strongest applicants
    - poorly performing employees tend to fired or quit

### Predictive validity studies

- Predictors collected first
- Criteria collected at a later period in time
- Sample = applicants
  - administer predictor measures to applicants
  - Do not use the predictors to make hiring decisions
- Advantages:
  - Uses an applicant sample
  - Avoids range restriction problems
- Disadvantage:
  - more complicated design
  - need to wait to collect criterion data

### Aggregating validity studies: Validity generalization (VG)

- VG involves aggregating the results from multiple validity studies to find a stable estimate of the validity of a given measure or construct for certain a criterion
- Validity estimates tend to vary from study to study, even when the exact same measures are used
- A more stable estimate of validity can be derived by aggregating(averaging) across validity studies
- studies with larger sample sizes are given more "weight" when computing the average, relative to studies with smaller sample sizes
- Variations in validity coefficients across studies is often largely due to sampling error and other statistical artifacts.

## Affect magnitude of observed validities

- Unreliability of the criterion
- range restriction

## Unreliability

- the correlation betwee na predictor and a criterion is attenuated by unreliability in the measures
- measurement error limits our ability to observe relationships between constructs
- Correction for attenuation formula

$$\hat{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{yy}}}$$

## Range restriction

- **Direct range restriction**: the measure bing validated was used to make hiring decisions
- range restriction leads to underestimation of the validity of predictors
- can apply a correction formula to estimate what the operational validity(i.e., the validity we'd expect in the applicant sample) from incumbent (i.e., current employee/student) data.

$$\hat{r}_{xy} = \frac{r_{xy}\frac{SD_u}{SD_r}}{\sqrt{1 - r_{xy}^2 + r_{xy}^2(\frac{SD_u}{SD_r})^2}}$$

  $\hat{r}_{xy}$ = corrected validity coefficient
  $r_{xy}$ = obvserved validity coefficient
  $SD_u$ standard deviation of predictor scores in unrestricted sample.
  $SD_r$ standard deviation of predictor scores in restricted sample.

- his formula demonstrates that the observed validity is reduced as a function of the **ratio of variability** in predictor scores across applicant and incumbent samples
- validities can be corrected for unreliable criteria and range restriction
- observed validities are typically underestimates of the operational validity