NATIONAL CENTER for INTERPROFESSIONAL PRACTICE and EDUCATION

August 3, 2017

FUNDAMENTALS OF IPECP MEASUREMENT II:

"Sources of Validity, Measurement Error, and Interpreting Your Results"

Connie C. Schmitz, PhD and Barbara Brandt, PhD

The National Center for Interprofessional Practice and Education is supported by a Health Resources and Services Administration Cooperative Agreement Award No. UE5HP25067. The National Center is also funded in part by the Josiah Macy Jr. Foundation, the Robert Wood Johnson Foundation, the Gordon and Betty Moore Foundation, The John A. Hartford Foundation and the University of Minnesota.

UNIVERSITY OF MINNESOTA

NATIONAL CENTER for INTERPROFESSIONAL PRACTICE and EDUCATION

JOINTLY ACCREDITED PROVIDER™
INTERPROFESSIONAL CONTINUING EDUCATION

In support of improving patient care, the National Center for Interprofessional Practice and Education is jointly accredited by the Accreditation Council for Continuing Medical Education (ACCME), the Accreditation Council for Pharmacy Education (ACPE), and the American Nurses Credentialing Center (ANCC), to provide continuing education for the healthcare team.

Physicians: The National Center for Interprofessional Practice and Education designates this live activity for a maximum of **1.5** *AMA PRA Category 1 Credits*™.

Physician Assistants: The American Academy of Physician Assistants (AAPA) accepts credit from organizations accredited by the ACCME.

Nurses: Participants will be awarded up to **1.5** contact hours of credit for attendance at this workshop.

Nurse Practitioners: The American Academy of Nurse Practitioners Certification Program (AANPCP) accepts credit from organizations accredited by the ACCME and ANCC.

Pharmacists: This activity is approved for **1.5** contact hours (.1 CEU) UAN: **0593-0000-17-007-H05-P**

The National Center for Interprofessional Practice and Education is supported by a Health Resources and Services Administration Cooperative Agreement Award No. UE5HP25067. The National Center is also funded in part by the Josiah Macy Jr. Foundation, the Robert Wood Johnson Foundation, the Gordon and Betty Moore Foundation, The John A. Hartford Foundation and the University of Minnesota.

UNIVERSITY OF MINNESOTA

NATIONAL CENTER for INTERPROFESSIONAL PRACTICE and EDUCATION

JOINTLY ACCREDITED PROVIDER™
INTERPROFESSIONAL CONTINUING EDUCATION

Disclosures

The National Center for Interprofessional Practice and Education has a conflict of interest policy that requires disclosure of financial interests or affiliations of organizations with a direct interest in the subject matter of the presentation.

Constance Schmitz and Barbara Brandt

do not have a vested interest in or affiliation with any corporate organization offering financial support or grant monies for this interprofessional continuing education activity, or any affiliation with an organization whose philosophy could potentially bias his/her presentation.

The National Center for Interprofessional Practice and Education is supported by a Health Resources and Services Administration Cooperative Agreement Award No. UE5HP25067. The National Center is also funded in part by the Josiah Macy Jr. Foundation, the Robert Wood Johnson Foundation, the Gordon and Betty Moore Foundation, The John A. Hartford Foundation and the University of Minnesota.

UNIVERSITY OF MINNESOTA

Connie C Schmitz, PhD

- Educational psychologist
  - Curriculum development, learner assessment, program evaluation, education research
- Consultant to the National Center
  - Measurement collection
- Experience
  - Academe, foundations and government agencies
- Scholar
  - Health and human service evaluation
  - 30 publications and 45 technical reports

Your Feedback is Important

- As stated in my July 12 webinar…
- NC is planning process for more programs on…
  - Measurement
  - Assessment
  - Evaluation
- Your feedback from today is critical
- Please complete survey (and quiz!)
- Answers will be mailed out

How to Be…

Assessment and Evaluation for IPECP

Better consumers of tools

More discerning readers

Better collaborators with measurement specialists

Fasten Your Seat Belts!

- Cover a lot of ground
- Have hand-outs available
- Paper and pen
- Pausing twice for Q+A

Replication Validation Study of ICCAS

FIPPC:

- Required course for 1,000 pre-licensure students at U of MN
- Nursing, medicine, dentistry, pharmacy, public health, vet med

The Interprofessional Collaborative Competency Attainment Survey (ICCAS): A replication validation study

Connie C. Schmitz, David M. Radosevich, Paul Jardine, Colla J. MacDonald, David Trumpower, and Douglas Archibald

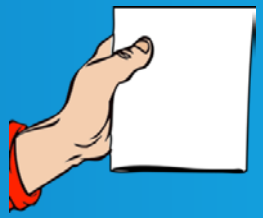
JOURNAL OF INTERPROFESSIONAL CARE 2017, VOL. 31, NO. 1, 28–34

ICCAS Content

20-item, self-assessment survey

1. Communication
2. Collaboration
3. Roles and responsibilities
4. Patient/family centered approach
5. Conflict management / resolution
6. Team functioning

Validation Study Abstract



see handout!

- “We appraised the content validity of the ICCAS for an IPE course and investigated its internal (factor) structure and concurrent validity.
- Self-assessed ratings were obtained from 785 pre-licensure, health professions students using a retrospective, pre-/post-design.
- Moderate to large effect sizes emerged for 16 of 20 items. Largest effects (1.01, 0.94) were for competencies emphasized in the course; the smallest effect (0.35) was for an area not directly taught.
- Positive correlations were seen between all individual item change scores and a separate item assessing overall change, and item-total correlations were moderate to strong.
- Exploratory factor analysis was used to understand the interrelationship of ICCAS items. Principal component analysis identified a single factor (Cronbach’s alpha = 0.96) accounting for 85% of the total variance.”

Questions about Abstract

- What does internal factor structure mean?
- What's an effect size?
- What's a change score?
- What does item-total correlation mean?
- Who is Cronbach?
- Is accounting for 85% total variance "good"?

Questions about Abstract

- Is this a good survey?
- Is it an accurate, fair assessment?
- Is it safe to make decisions based on these scores?

These are questions of validity

13

Agenda

1. Foundation concepts

(pause for questions)

2. Validity evidence

(pause for questions)

3. Rules of thumb
4. Revisit example
5. Wrap up

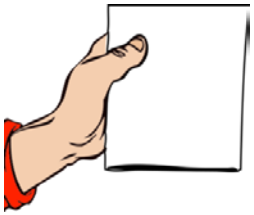
1. Foundation Concepts

- **Definition of validity**
- Variance
- Measurement error
- Reliability

Validity is More than Skin Deep

- Characteristics of “good tools”
  - Look, feel, tone
  - Basic construction, formatting
  - Clear instructions, wording, scoring
  - Reasonable scenarios, tasks
  - Absence of obvious bias

- “Face” validity ≠ validity

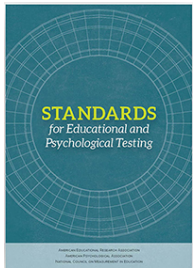


See handout!

Validity Defined

- Trustworthiness

- Integrated judgment about the degree to which evidence collected about a test supports the way we interpret and use its scores*



*AERA, APA, NCME

Standards for Educational and Psychological Testing, 2014

Validity is not Inherent to the Tool

- We don’t validate instruments, but our interpretation and use of scores
- Validity is a matter of degree
- Validity data are sensitive to context
- Validity data fluctuate

The Process of Validation

1. Develop “claims” about what is being measured and what scores mean

2. Collect data (evidence) to test these claims

3. Make a judgment, based on data

4. Replicate, look for consistent validity data across place and time

Example

"IPCP Leadership Scale"

Example: “IPCP Leadership Scale”

- Claims:
  - These are the right behaviors
  - People observing leaders can agree and score alike
  - Leadership scores correlate with ratings from practice

- Evidence:
  - Literature review
  - Scores from trained raters agree
  - Trained rater scores correlate with 360 degree evals

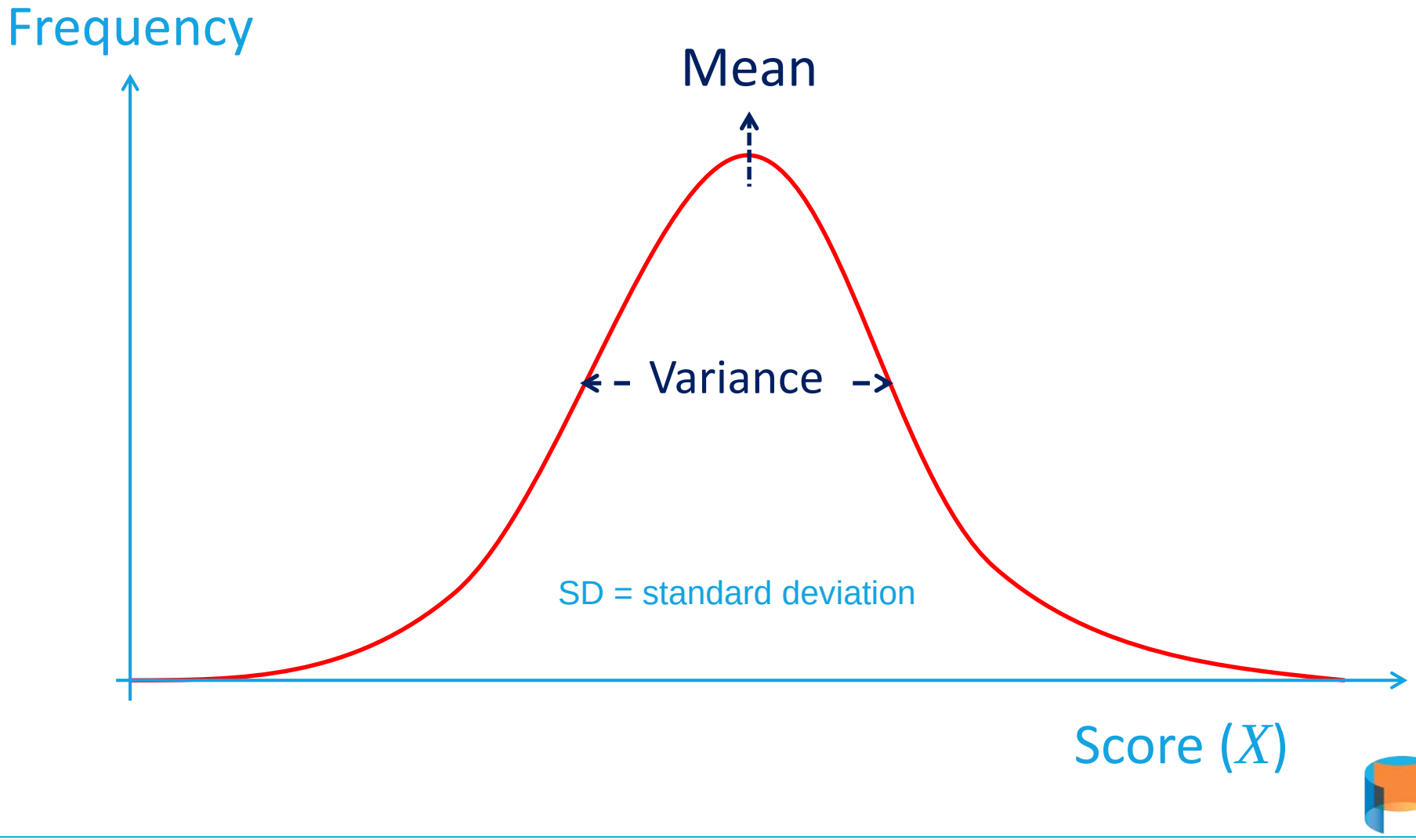
1. Foundation Concepts

- Definition of validity
- **Variance**
- **Measurement error**
- Reliability

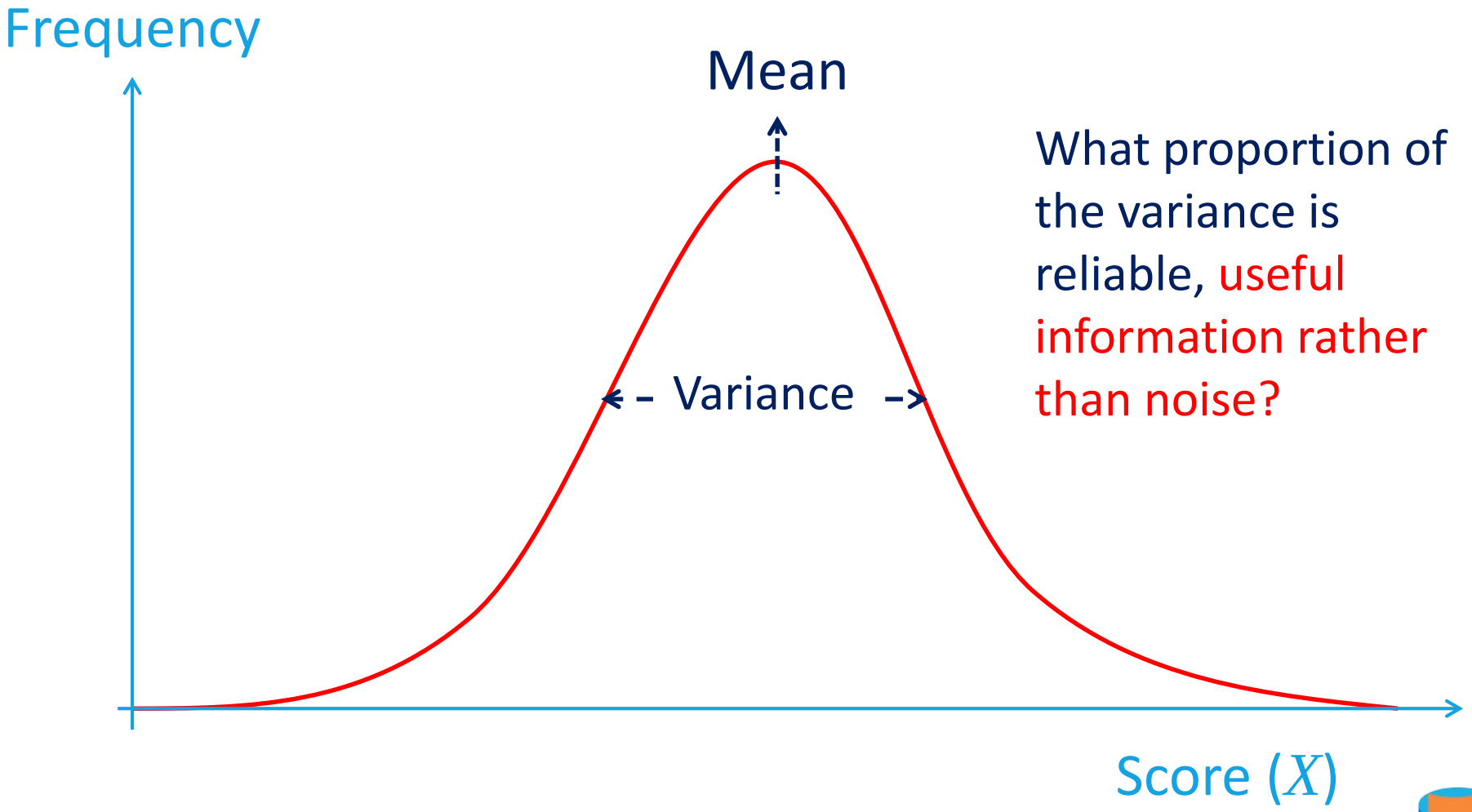
Variance

- Data vary!!!
- Types of data
  - Multiple choice, checkbox, yes/no (categories)
  - Scales
- Summary statistics
  - total number correct
  - total number per category
  - mean scale score
  - mean overall scale score
  - mean change score (post-score – pre-score)
  - standardized scores

Normal Curve

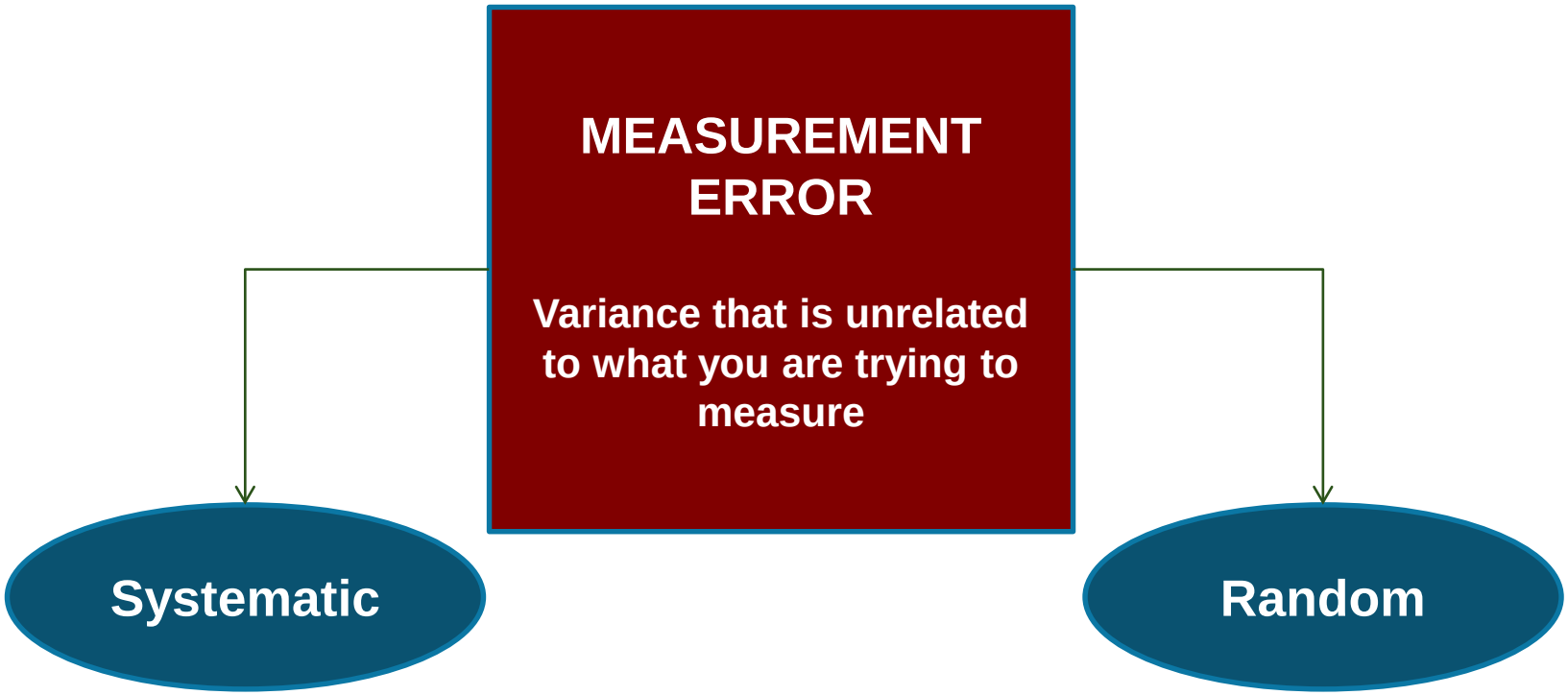


Q: Meaningful Variance?



What proportion of the variance is reliable, useful information rather than noise?

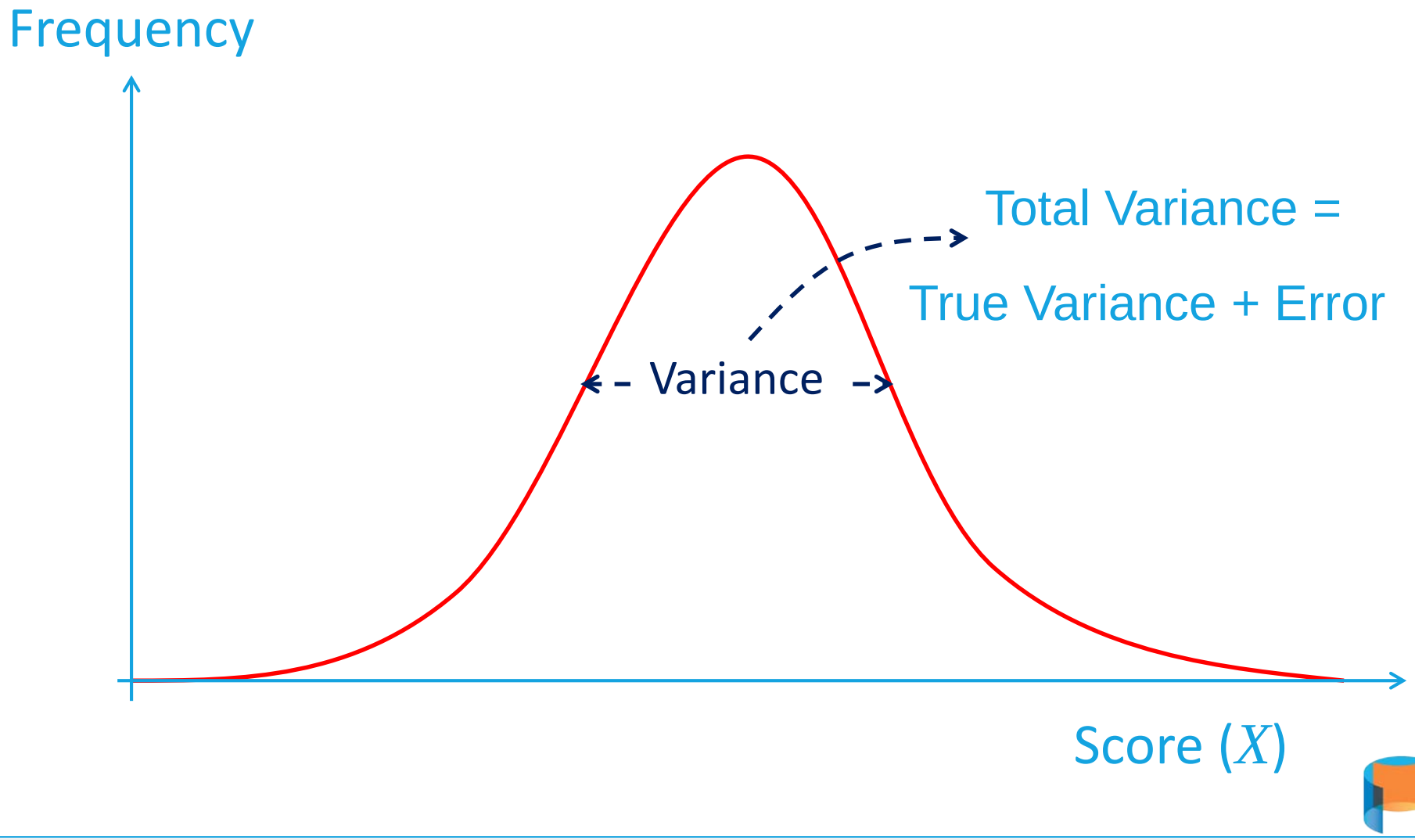
Measurement Error



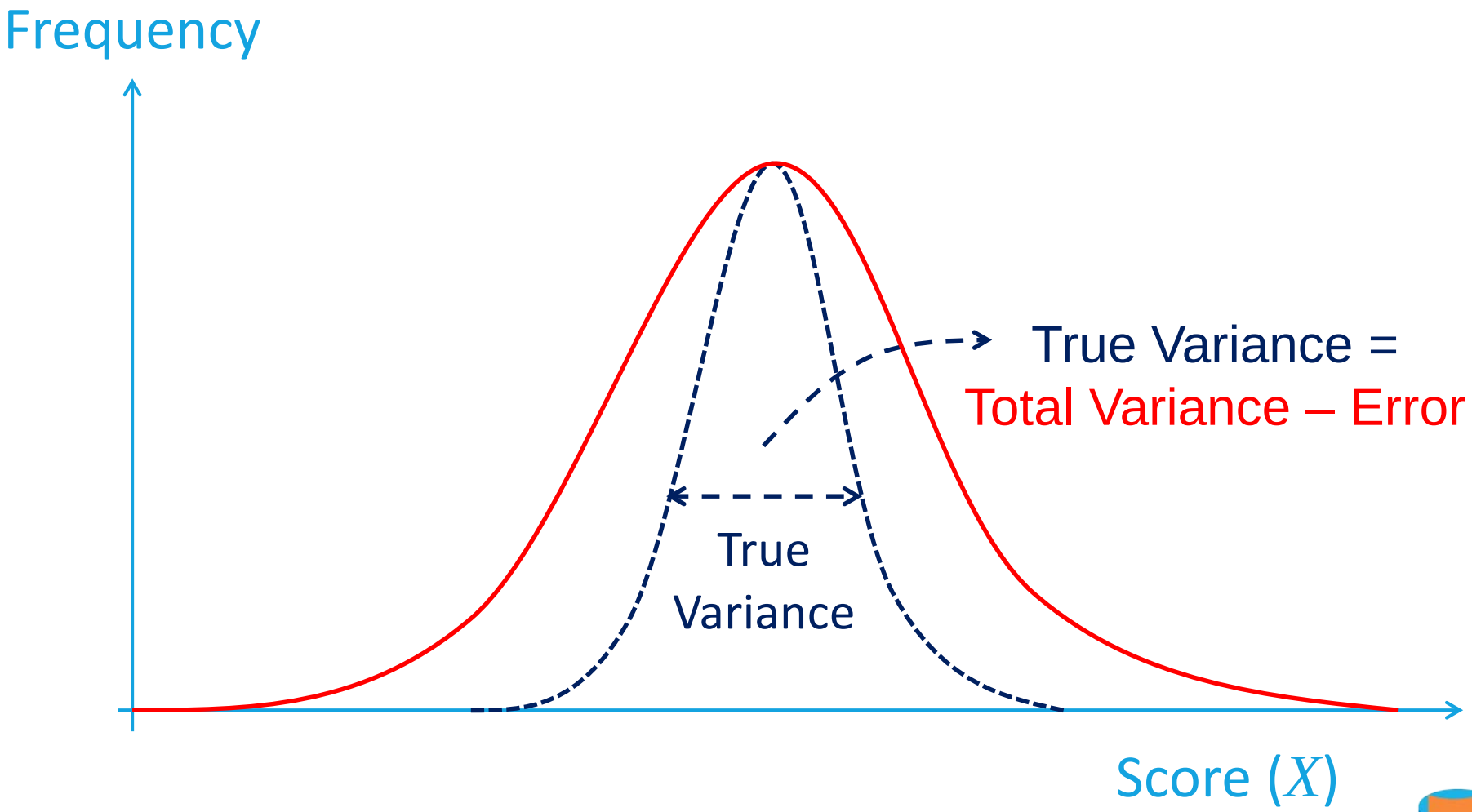
Across all subjects
- ❑ Problems with instrument
- ❑ Problems with administration
- ❑ Problems with incentives

Varies with subject
- ❑ Chance fluctuations
- ❑ Different backgrounds
- ❑ Motivation, fatigue

Total Variance



True Variance



1. Foundation Concepts

- Definition of validity
- Variance
- Measurement error
- **Reliability**

Reliability

- "Precision" of measurement
  - Proportion of true variance is high
  - Measurement error is low
- "Consistency" of scores
  - Items measuring the same construct generate similar responses
- "Reproducibility" of scores
  - Individuals taking the same test twice will score the same
  - Different raters assessing the same subjects will score them alike

Common Reliability Statistics

- Reliability coefficients (estimates): r = 0.00 - 1.00
  - Internal consistency (Cronbach's alpha, α)
  - Intra-class correlation (ICC)
  - Inter-rater reliability (r)
  - Inter-rater agreement (Cohen's *kappa*)
- Other techniques
  - Standard Error of the Measurement (SEM)
  - Confidence intervals (probability) (CI)

How Reliable?

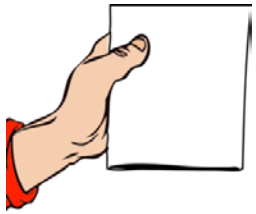
- Reliability coefficients
  - .90 for high stakes
  - .80 - .89 for moderate
  - .70 - .79 for formative, low stakes

Downing SM, Reliability: On the reproducibility of assessment data. *Med Educ*, 2004; 38:1006-12

Why is Reliability Important?

- Integral to validity!

- For more information…..



Questions?

- Definition of validity
- Variance
- Measurement error
- Reliability

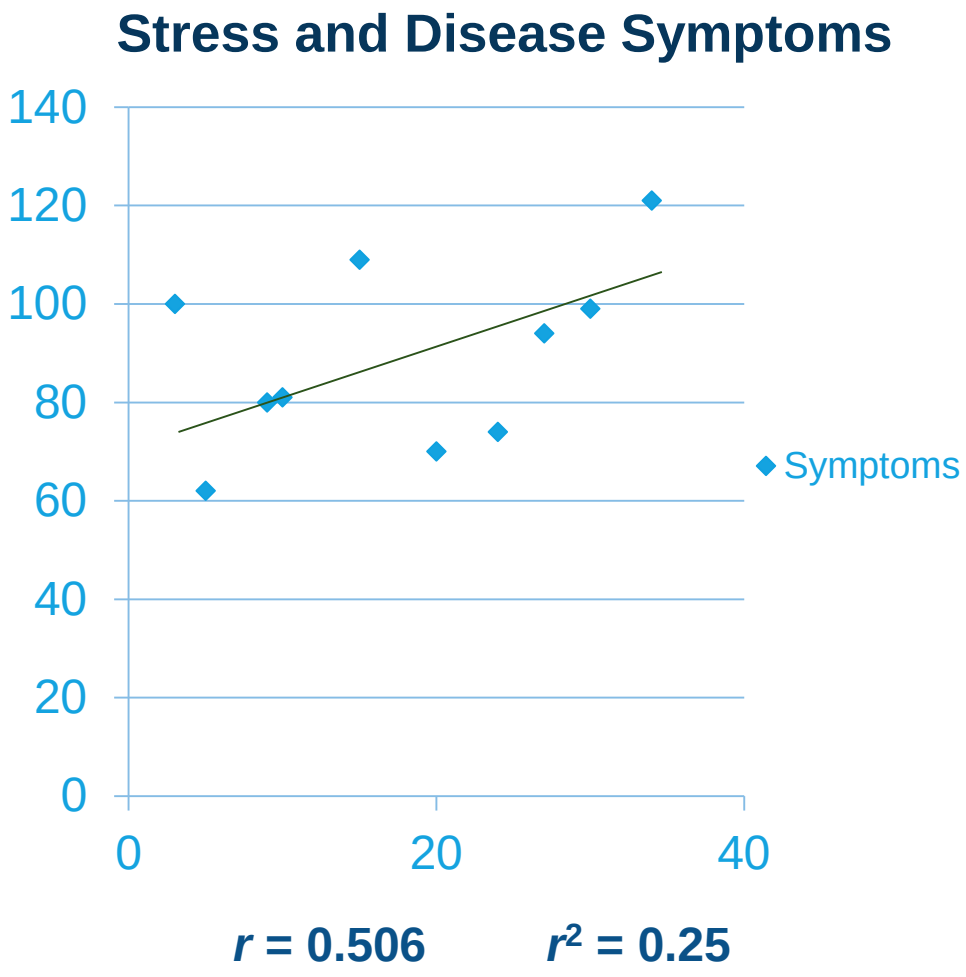
2. Validity Evidence

- **Nature of validity evidence**
- Sources of validity evidence
- Where can validity data be found?

The Nature of Validity Evidence

- Validity evidence = qualitative / quantitative
- Validity data = statistics = “estimates”
  - Correlation coefficients (*r*)
  - Factor loading (“loading” 0.00)
  - % of variance accounted for (%)
  - Effect size (Cohen’s *d*)

Correlation



Subject	Stress Scores	Symptom Scores
1	34	121
2	30	99
3	27	94
4	24	74
5	20	70
6	15	109
7	10	81
8	9	80
9	5	62
10	3	100

Correlation Coefficients

- Basis for many statistical techniques used
- Zero to +1, Zero to -1 (r = 0.XXX)
- Squared correlation (r^2 = 0.XXX)

Example from previous slide:

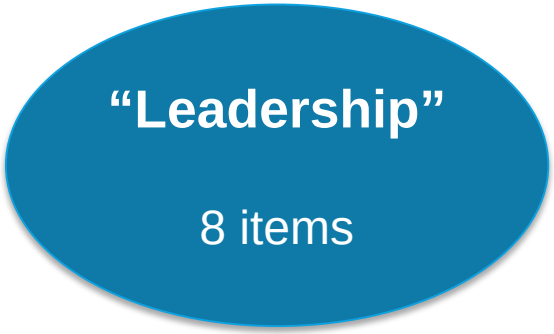
r = 0.506 **r^2 = 0.25** ⟶ 25% of the total variance in the data set relates to the relationship between these two variables

Factors

- Extent to which item scores correlate together according to constructs (factors)
  - “leadership”
  - “teamwork”
- Exploratory vs. confirmatory
- Helps us understand instrument content, and whether it is measuring what we think it is measuring

Factor Loading: Example

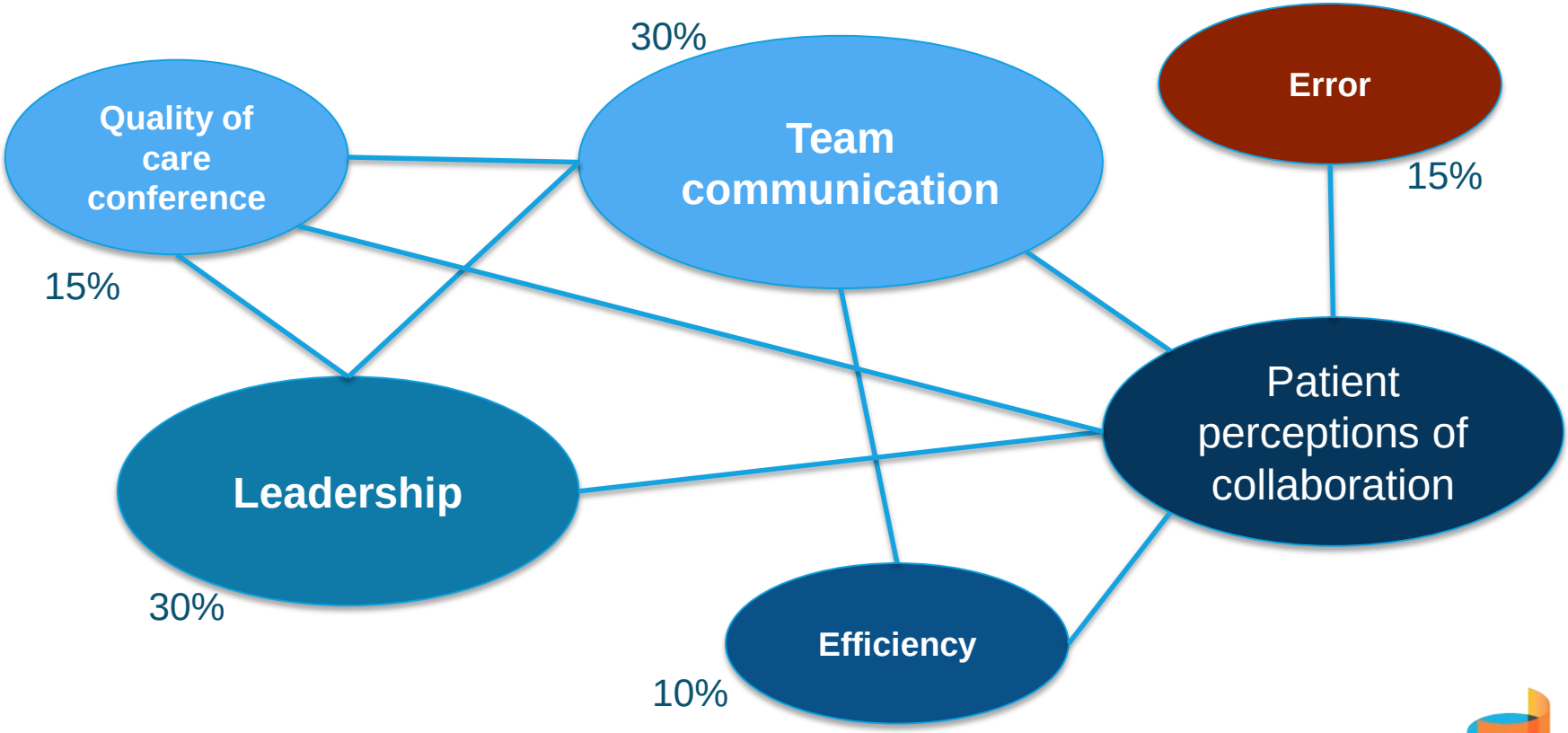
- Degree to which item scores correlate to a common factor
- Item loadings (>0.40)



Leadership Items	Factor Loadings
1. Puts "team" ahead of "self"	0.89
2. Shares information with others	0.78
3. Invites feedback	0.83
4. Respects team members	0.82
5. Solves problems collaboratively	0.81
6. Can set agendas	0.50
7. Is a friendly colleague	0.32
8. Has excellent patient rapport	0.15

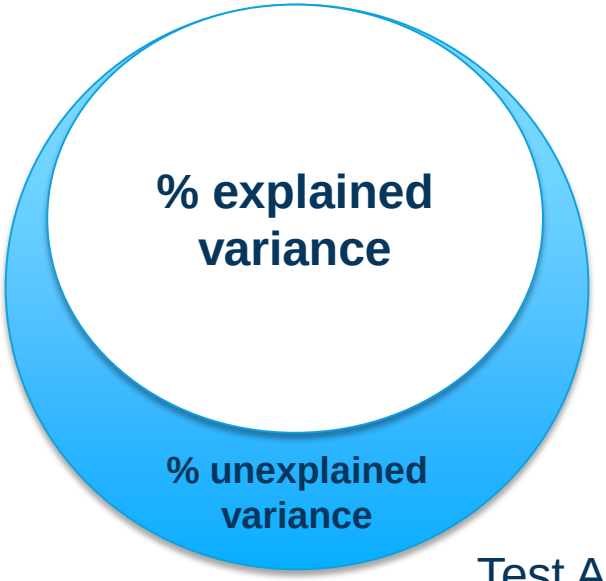
Explaining Variance: Example

- How much of variance is due to specific factors?

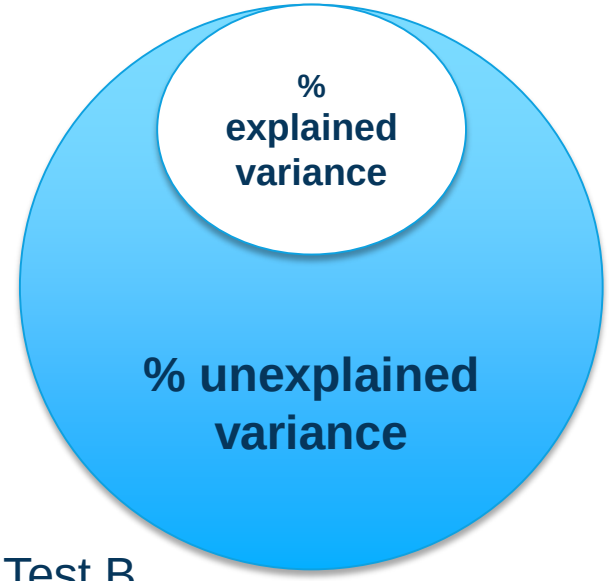


Importance of Explained Variance

- Integral to validity!
- Eigenvalues over 1.0
- Percentage



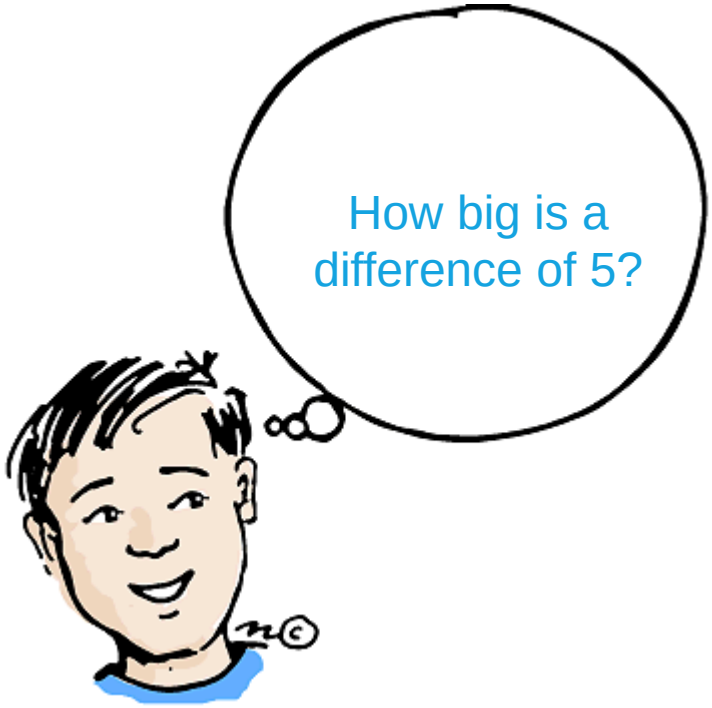
Test A



Test B

Effect Size

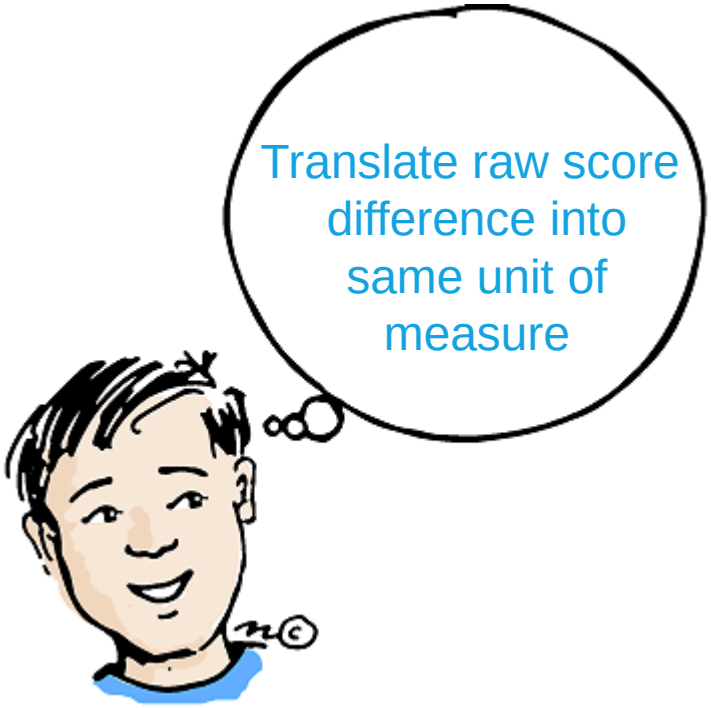
- Magnitude of change or difference
  - Pre-post
  - Between groups



- Raw difference
  - post-score (x=20) minus pre-score (x=15) = 5

Standardized Effect Size

- Standardized
  - Post-score (20) – pre-score (5) = 5
  - 5 divided by the SD (4.5) = **1.11**



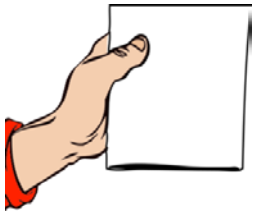
- Interpreting effect sizes
  - Large = ≥0.80
  - Moderate = between 0.50 and 0.79
  - Small = <0.50

2. Validity Evidence

- Nature of validity evidence
- **Sources of validity evidence**
  - Claims
  - Types of evidence
- Where can validity data be found?

Sources of Validity Evidence*

1. Content
2. Response process
3. Internal structure
4. Relationship between scores and other variables
5. Consequences of testing

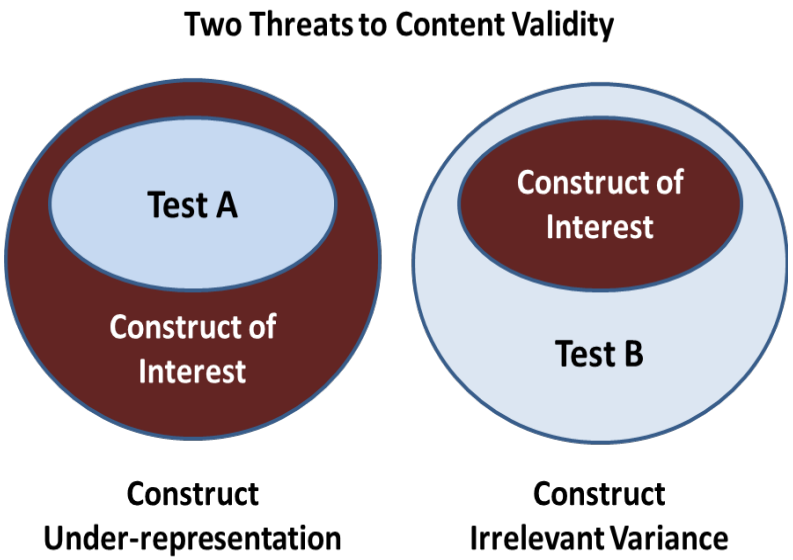


IPECP examples

*AERA, APA, NCME, Standards for Educational and Psychological Testing, 2014

Claim 1: Content

- Problem: Finite space + time: not everything can be measured
- Fact: Items (tasks, cases) are a “sample”
- Claim: Content = good sample; represents the thing you are trying to measure



Content: Types of Evidence

- Literature review
- Blueprints
- Expert review
- Consensus panels

"Provenance"

Justification for content

Claim 2: Response Process

- Problem: Tests are not real life
- Fact: Items, tasks, cases may/may not stimulate or allow a learner to respond authentically
- Claim: Assessment provides fair opportunity to respond

Response Process: Types of Evidence

- Cognitive interviews
- Observation
- Examining forms for blanks, anomalies
- Rater debriefing
- Inter-rater reliability
- Inter-rater agreement

Authentic response

Knowledgeable scoring

Claim 3: Internal Structure

- Problem: It is hard to write good items
- Fact: Some items may be flawed, unreliable, irrelevant, or redundant
- Claims:
  - measurement error is low
  - reliability is high
  - factors are independent, stable
  - variance can be explained

Internal Structure: Types of Evidence

- Internal consistency (Cronbach's alpha)
- Scale reliability
- Inter-class correlation
- Item-total correlation
- Generalizability
- Exploratory, confirmatory factor analysis
- Items fit model

Reliable

Accurate

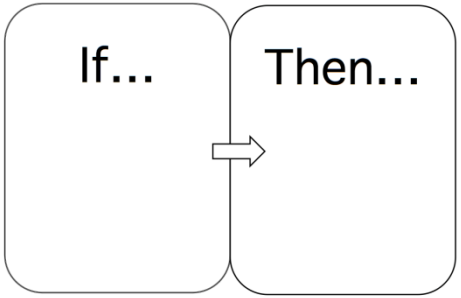
Claim 4: Relationships with Other Variables

- Problem: What is the value of a test result? How do we interpret, use it?

Signs + symptoms

- Fact: Scores gain meaning if we know how they relate to other things.

Claims about scores:



Relationships: Types of Evidence

- Convergent (concurrent): scores correlate with similar measures, variables
- Divergent: scores do *not* correlate with unrelated measures, variables
- Discriminant: scores differentiate among groups
- Predictive: scores correlate with outcomes

Meaning

Utility

Claim 5: Consequences

- Problem: Good assessment is hard
- Fact: Assessment can lead to unintended consequences
- Claim: Benefits outweigh the costs

Benefits

- Improved planning
- Student learning
- Better prediction
- Fairer decisions, policies

Costs

- Resources
- Misuse of scores
- Political fallout
- Replication failure

Consequences: Types of Evidence

- Stakeholder feedback
- Actual resource use
- Evidence of misuse
- Decision accuracy
- Effects on desired outcomes

Benefits

outweigh

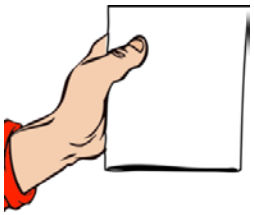
Costs

Questions?

- Nature of validity evidence
- Sources of validity evidence

Where Can Evidence be Found?

- Manuals, guides
- Reference books
- National Center's measurement instrument collection
- Journal articles

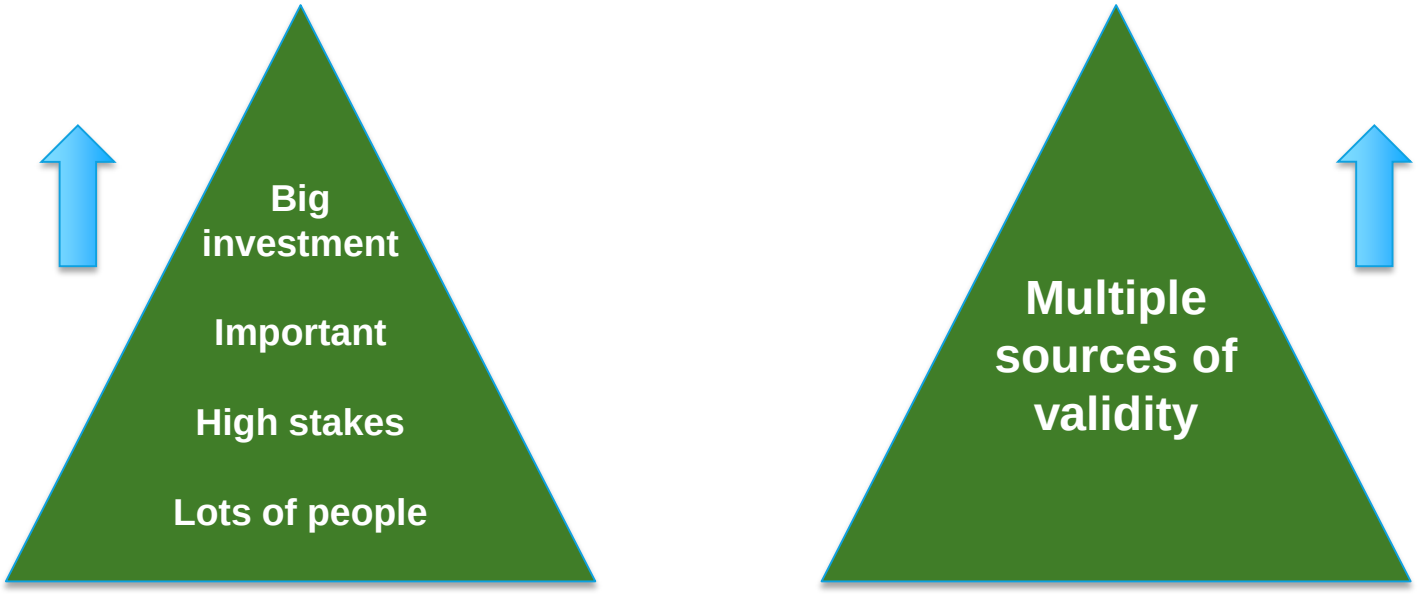


Where are we in the Agenda?

1. Foundation concepts
2. Validity evidence
3. **Rules of thumb**
4. **Revisit example**
5. Wrap up

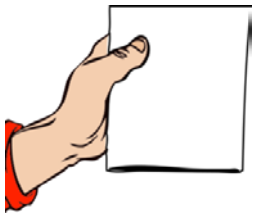
3. Rules of Thumb

- Collecting validity data = time, money, expertise
- How much validity data is “enough?”
- Integrated judgment: quality of data, purpose, context



Rules of Thumb: How High?

- Pattern across data
- See handout



Connie's top 10 Statistics

- Reliability coefficients
- SEM
- Confidence intervals
- Factor loadings
- Eigenvalues
- Validity coefficients
- Factor loadings
- % of variance explained
- Effect sizes
- Item-total correlations

4. Return to ICCAS Example

Validity claims we tested

- ✓ Content
- ✓ Internal structure
- ✓ Relationship to other variables (concurrent validity)

Validity claims we did not test

- Response process
- Consequential

ICCAS Content Validity

- Before administration
  - Reviewed original study
  - Items aligned with FIPCC course objectives, IPEC core competencies
- After administration
  - Self-reported gains = “change scores”
  - Degree of change = “effect sizes”
  - Effect sizes were high in areas covered by course, low in areas not covered

ICCAS Change Scores

ICCAS Construct	ICCAS Item	UM Students (n = 785)	
		Cohen's *d*	Difference
Communication	• Promote effective communication among IP members	0.72	Moderate
	• Actively listen to IP team members' ideas, concerns	0.51	Moderate
	• Express my ideas and concerns without being judgmental	0.54	Moderate
	• Provide constructive feedback to IP team members	0.52	Moderate
	• Express my ideas clearly and precisely	0.39	Small
Collaboration	• Seek out IP team members to address issues	0.78	Moderate
	• Work closely with IP team members to enhance care	0.72	Moderate
	• Learn from IP team members to enhance care	0.94	Large
Roles and Responsibilities	• Identify and describe my abilities and contributions to the IP team	0.72	Moderate
	• Be accountable for my contributions to the IP team	0.43	Small
	• Understand the abilities and contributions of IP team members	1.01	Large
	• Recognize how others' skills and knowledge complement my own	0.98	Large
Patient Centered Care	• Use an IP team approach with the patient to assess health	0.74	Moderate
	• Use an IP team approach with the patient to provide whole person care	0.69	Moderate
	• Include the patient / family in decision making	0.35	Small
Conflict Management, Team Functioning	• Actively listen to the perspective of IP team members	0.55	Moderate
	• Take into account the ideas of IP team members	0.60	Moderate
	• Address team conflict in a respectful manner	0.43	Small
	• Develop an effective care plan with IP team members	0.75	Moderate
	• Negotiate responsibilities within overlapping scopes of practice	0.79	Moderate

ICCAS Internal Structure

- Factor structure
  - Single underlying factor (not 6)
  - 85% of variance explained is “good”
  - Cronbach’s alpha of 0.96 is “excellent”
  - Item-total correlations = “moderate” to “strong”

ICCAS External Variables

- External variables
  - Correlation between individual items and change in ability to collaborate interprofessionally

20 effect sizes

Pre-post change in each of 20 behaviors

20 correlations

18 items r = ≥ .40
02 items r = < .40

1 global item score

Change in overall ability to collaborate

Change in Overall Ability Scale
1 = much better now; 2 = somewhat better now;
3 = about the same; 4 = somewhat worse now;
5 = much worse now

ICCAS: Summary

- Content validity = high
  - Sensitive to FIPCC content
- Internal structure = unexpected, but good
  - Single factor explained large proportion of variance
  - High internal consistency
- Relation to other variables = good
  - Behaviors align with overall change in collaborative ability

ICCAS: Conclusions

- ✓ Is this a good survey?
- ✓ Is it an accurate, fair assessment?
- ✓ Is it safe to make decisions based on these scores?

5. Wrap Up

- Validity is more than skin deep
  - Characteristics of tool construction matter
  - Support lies in validity data
- Appreciate the difficulty of determining validity
  - Be skeptical of “reliable and valid tool”
  - Look for stable validity data across settings
  - Collect local response process validity
  - Limit error through standardization, quality control
  - Collect local reliability data
- Work with measurement professionals

In Closing

Okay to unbuckle…

Further reading

[nexusipe.org/advancing/assessment-evaluation-start]

- Measurement Primer
- Practical Guide Series

Questions

Acknowledgements

- Michael Cullen, PhD
  - Director of Evaluation
  - University of Minnesota Medical School
- David Radosevich, PhD
  - University of Minnesota, Epidemiologist / Biostatistician (retired)
  - Consultant to the National Center
- Jane Miller, PhD
  - Director of Simulation
  - University of Minnesota Academic Health Center
- Yoon Soo Park, PhD
  - Associate Professor
  - University of Illinois at Chicago, Dept Medical Education

NATIONAL CENTER for INTERPROFESSIONAL PRACTICE and EDUCATION

The National Center for Interprofessional Practice and Education is supported by a Health Resources and Services Administration Cooperative Agreement Award No. UE5HP25067. The National Center is also funded in part by the Josiah Macy Jr. Foundation, the Robert Wood Johnson Foundation, the Gordon and Betty Moore Foundation, The John A. Hartford Foundation and the University of Minnesota.

UNIVERSITY OF MINNESOTA