

Designing Surveys, Questionnaires and Other Instruments

An Introduction

Georgios D. Sideridis, Ph.D.

Boston Children's Hospital, ICCTR

Harvard Medical School

georgios.sideridis@childrens.harvard.edu



Children's Hospital Boston

Institutional Centers for Clinical and Translational Research

[Donate](#)

[Home](#)[Research Resources](#)[Education](#)[Funding](#)[For Patients](#)[About Us](#)[Search](#)



Institutional Centers for Clinical and Translational Research (ICCTR)

[Contact Us](#)[Request Services](#)

[About Us](#)[Biostatistics & Research Design](#)[Experimental Therapeutics/Interventional Trials](#)[Population Health Sciences](#)[Translational Research Program](#)[News & Announcements](#)[Contact Us](#)

ICCTR | About Us | Biostatistics & Research Design (BARD)



Learn more about the ICCTR Biostatistics and Research Design (BARD) Center, including:

- [Mission](#)
- [Services](#)
- [Education](#)
- [Careers](#)
- [FAQs](#)
- [Team](#)

Mission

The mission of the BARD Center is to promote excellence in the design and analysis of research studies at Boston Children's Hospital. The ICCTR BARD Center provides biostatistical and methodological expertise, as well as scientific leadership through collaborative relationships with Boston Children's investigators. The Center also provides educational offerings in a variety of areas.



Children's Hospital Boston

Organization

- Describe the principles of survey development
- Present common pitfalls with item wording and scaling,
- And,
- Discuss issues of reliability and validity
- +
- Take a brief quiz?



Items from the Group

Scaling System	Wording of Items	Reliability and Validity
*Knowing which "rules" to follow! For example, I have heard conflicting guidance about whether to formulate survey items as statements or questions. I have also heard conflicting advice about whether to use "strongly disagree, disagree, neutral, agree, strongly agree" or to customize the answers for each question. *Anchoring the response options *creating response categories for the audience being surveyed, i.e., it is always a team sport. *validation, wording of questions, appropriate Likert scale	*Clear, concise wording of questions *ensuring no bias is in questions *question writing *Creating questions that are free from bias *Not leading on/biasing responses *Getting the questions right so they are interpreted by all users the same way. *Formulating clear questions	*Need better understanding of psychometrics *Ensuring I have validity and reliability *Assuring validity *Validating



Should I use an existing survey or try to develop my own?

- Please don't....
- Unfortunately most of the time you have to because there is no suitable instrument
- Better to modify an existing one
- What if the existing one has poor psychometric properties or no documented properties?

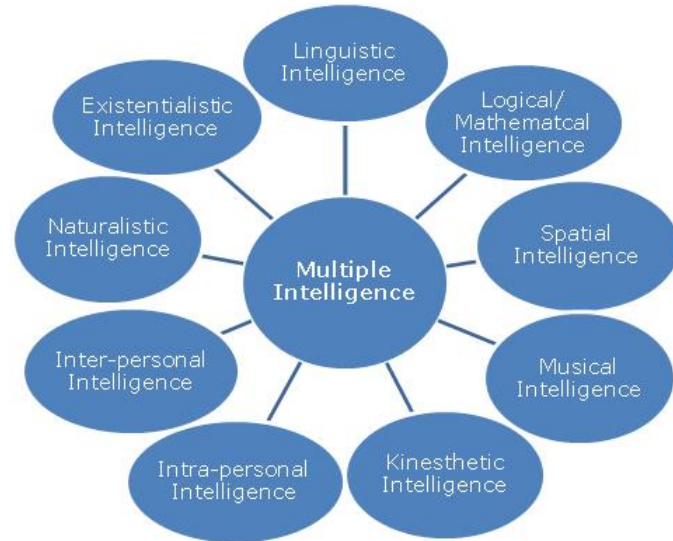
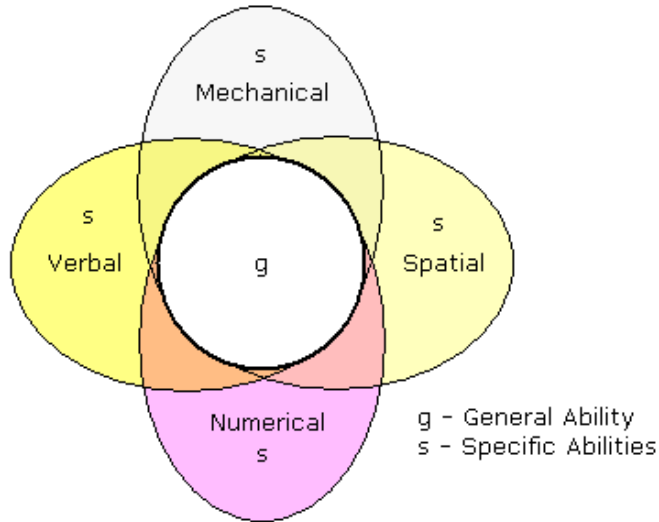


The Principles of Survey Development Work.....

-except when they don't!
- There are always instances and specifics in a study that render these suggestions meaningless:
- E.g., a Likert scaling system (Agree-Disagree) is likely more sensitive compared to a dichotomous yes-no scaling



What's the first and most important step in survey development? Theory, why?

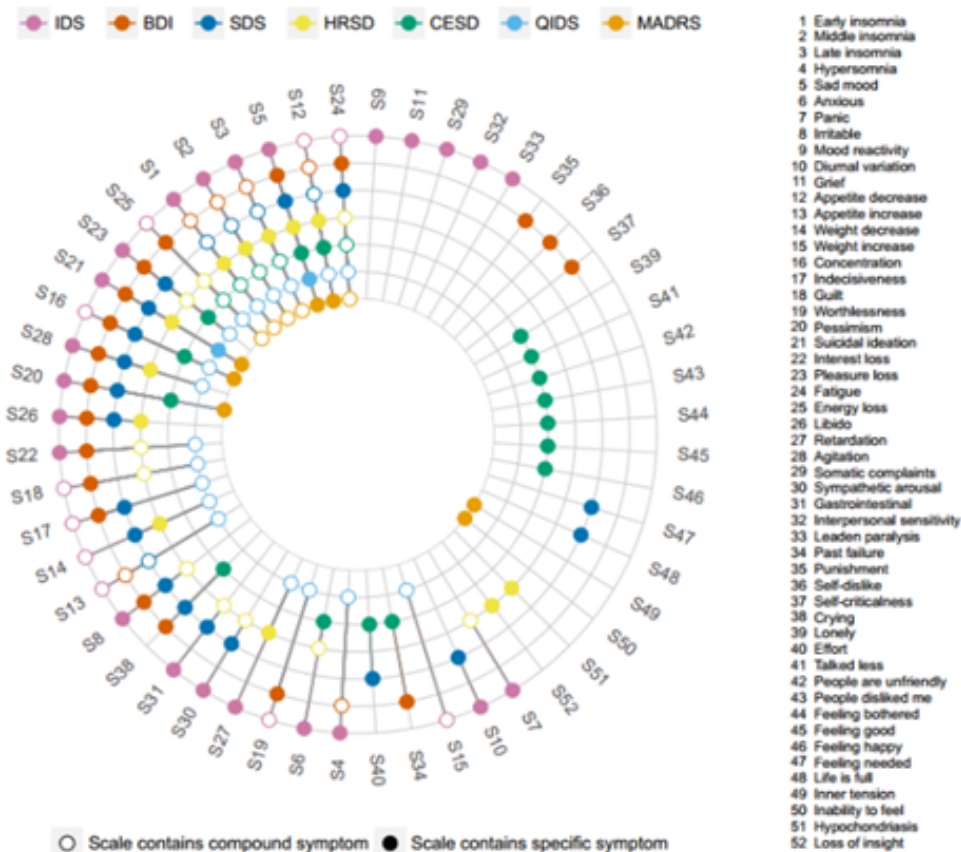


Goal constructs were developed based on intrinsic-extrinsic
Motivational theory and the dichotomy developed since then to:
Mastery-performance
Trichotomous theory
2X2



Other important steps in survey development?

- Constructs: Mostly unobservable: e.g., motivated behavior, being wealthy, helpless, bullying, neuroticism, etc.
- Operational Definitions (e.g., SES)
- Operational definition of aggressive behavior: “every instance that the child hits another child”
- We are being judged by both our theory and operational definitions



Content

Theory

Past research

Your insights

Contact experts?



What's in an operational definition?

- An operational definition is a way to define a behavior in simple, observable, and measurable terms.



Operational Definition: Physical Aggression

- Every behavior and instance in which a person intends to cause physical harm to another person. Items:
- Hitting
- Kicking
- Head banging
- Punching
- Scratching
- Pinching
- Biting
- Pushing
- Throwing objects
- Hair pulling
- Spitting
- Slapping
- Cutting



Any type of forcing ones self or objects toward another person

Children's Hospital Boston

So where does the content of the items come from?

- Theory
- Operational definitions
- Empirical findings
- Other instruments
- Informal observations
- Example topic: **Satisfaction with Marriage**
- Research has shown that married men live longer compared to single men
- To inform the measurement of satisfaction with marriage (item development) a series of interviews took place →



Stages of Survey/Instrument Development

- Selection of topic (based on need?)
- Selection of item type and format (e.g., multiple choice, alternate form, rating scale, forced choice, checklist)
- Selection of item presentation type (responses below the item? Next to the item?)
- Arrange coding schemes (reversed coding?)
- All decisions should be literature-based
- Pilot test
- Psychometric check before moving on

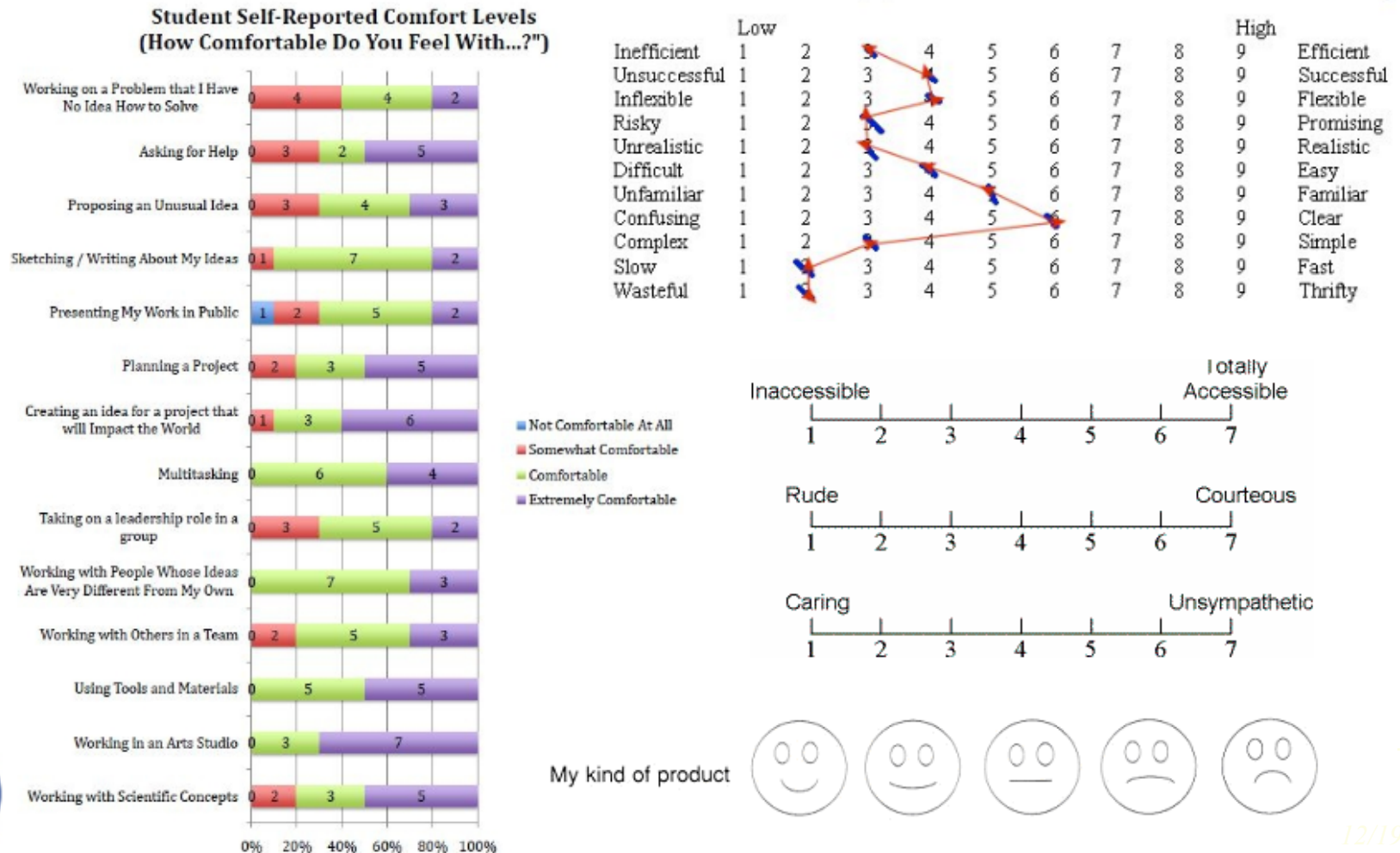


Item Wording and Item Scaling



Scaling of Surveys and Other Instruments

- Which scaling option scheme?
- Can we evaluate its efficiency?



Multiple Choice

- Selection of one response among several others
- **Example:**
 - The opposite meaning of “corruption” is:
 1. interruption
 2. construction
 3. diversion
 4. empathy
 5. honesty



Rating Scale

- We have many more, alternative, forced choice, multiple-choice but the most common is the rating scale
- The choices are within a continuum from $-\infty$ to $+\infty$
- **Example**

I like going to places where nobody has gone before	Strongly Disagree	Disagree	Nor agree nor disagree	Agree	Strongly agree
---	----------------------	----------	---------------------------	-------	-------------------

*Health care is under-financed: how does the mid response fit?



Rating Scale

Level of Agreement

- 1 – Strongly disagree
- 2 – Disagree
- 3 – Somewhat disagree
- 4 – Somewhat agree
- 5 – Agree
- 6 – Strongly agree

Frequency

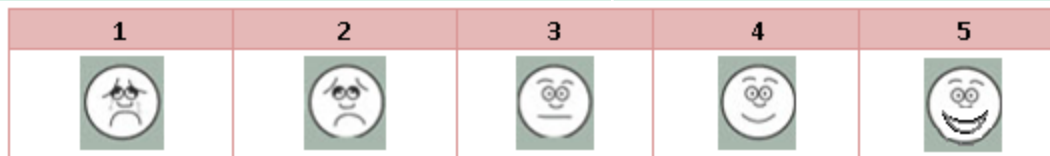
- 1 – Never
- 2 – Rarely, in less than 10% of the time
- 3 – Occasionally, in about 30% of the time
- 4 – Sometimes, in about 50% of the time
- 5 – Frequently, in about 70% of the time
- 6 – Usually, in about 90% of the time
- 7 – Always

Level of Usefulness

- 1 – not at all useful
- 2 – slightly useful
- 3 – somewhat useful
- 4 – moderately useful
- 5 – very useful
- 6 – extremely useful

Level of Satisfaction

- 1 – Completely dissatisfied
- 2 – Mostly dissatisfied
- 3 – Somewhat dissatisfied
- 4 – Somewhat satisfied
- 5 – Mostly satisfied
- 6 – Completely satisfied



Likert-Type Item Scaling Options

Likert-Type Scale Response Anchors

Citation:

Vagias, Wade M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

Level of Acceptability

- 1 – Totally unacceptable
- 2 – Unacceptable
- 3 – Slightly unacceptable
- 4 – Neutral
- 5 – Slightly acceptable
- 6 – Acceptable
- 7 – Perfectly Acceptable

Level of Appropriateness

- 1 – Absolutely inappropriate
- 2 – Inappropriate
- 3 – Slightly inappropriate
- 4 – Neutral
- 5 – Slightly appropriate
- 6 – Appropriate
- 7 – Absolutely appropriate

Level of Importance

- 1 – Not at all important
- 2 – Low importance
- 3 – Slightly important
- 4 – Neutral
- 5 – Moderately important
- 6 – Very important
- 7 – Extremely important

Level of Agreement

- 1 – Strongly disagree
- 2 – Disagree
- 3 – Somewhat disagree
- 4 – Neither agree or disagree
- 5 – Somewhat agree
- 6 – Agree
- 7 – Strongly agree

Knowledge of Action

- 1 – Never true
- 2 – Rarely true
- 3 – Sometimes but infrequently true
- 4 – Neutral
- 5 – Sometimes true
- 6 – Usually true
- 7 – Always true

Reflect Me?

- 1 – Very untrue of me
- 2 – Untrue of me
- 3 – Somewhat untrue of me
- 4 – Neutral
- 5 – Somewhat true of me
- 6 – True of me
- 7 – Very true of me

My beliefs

- 1 – Very untrue of what I believe
- 2 – Untrue of what I believe
- 3 – Somewhat untrue of what I believe
- 4 – Neutral
- 5 – Somewhat true of what I believe
- 6 – True of what I believe
- 7 – Very true of what I believe

Priority:

- 1 – Not a priority
- 2 – Low priority
- 3 – Somewhat priority
- 4 – Neutral
- 5 – Moderate Priority
- 6 – High priority
- 7 – Essential priority

Level of Concern

- 1 – not at all concerned
- 2 – Slightly concerned
- 3 – Somewhat concerned
- 4 – Moderately concerned
- 5 – Extremely concerned

Priority Level

- 1 – Not a priority
- 2 – Low priority
- 3 – Medium priority
- 4 – High priority
- 5 – Essential

Level of Problem

- 1 – Not at all a problem
- 2 – Minor problem
- 3 – Moderate problem
- 4 – Serious problem

Affect on X

- 1 – No affect
- 2 – Minor affect
- 3 – Neutral
- 4 – Moderate affect
- 5 – Major affect

Level of Consideration

- 1 – Would not consider
- 2 – Might or might not consider
- 3 – Definitely consider

Level of Support/Opposition

- 1 – Strongly oppose
- 2 – Somewhat oppose
- 3 – neutral
- 4 – Somewhat favor
- 5 – Strongly favor

Level of Probability

- 1 – Not probable
- 2 – Somewhat improbable
- 3 – Neutral
- 4 – Somewhat probable
- 5 – Very probable

Level of Agreement

- 1 – Strongly disagree
- 2 – Disagree
- 3 – Neither agree or disagree
- 4 – Agree
- 5 – Strongly agree

Level of Desirability

- 1 – Very undesirable
- 2 – Undesirable
- 3 – neutral
- 4 – Desirable
- 5 – Very desirable

Level of Participation

- 1 – No, and not considered
- 2 – No, but considered
- 3 – Yes

Frequency – 5 point

- 1 – Never
- 2 – Rarely
- 3 – Sometimes
- 4 – Often
- 5 – Always

Frequency

- 1 – Never
- 2 – Rarely
- 3 – Occasionally
- 4 – A moderate amount
- 5 – A great deal

Frequency of Use

- 1 – Never
- 2 – Almost never
- 3 – Occasionally/Sometimes
- 4 – Almost every time
- 5 – Every time



Children's

Things to Avoid...

- Avoid asking two things at the same time
e.g., how much do you like soccer and basketball?
Not at all <-----> Very Much so
- Avoid vagueness
e.g., Have you ever had experience with data collection?



Things to be careful with...

- Simple language (5th grade)

- **Wrong:**

A variable interval schedule of reinforcement is resistant to extinction

1. SD
2. D
3. A
4. SA

- **Right:**

When a person does not expect reinforcement regularly his/her behavior will likely persist longer

1. SD
2. D
3. A
4. SA



Things to be careful with...

- Avoid relations between items

Wrong

- The motive to achieve:
 - 1. relates to affect
 - 2. relates with self-regulation
 - 3. relates with self-efficacy
 - 4. all the above depend on achievement levels
- So, its relationship to actual achievement:
 - 1. is null
 - 2. is positive
 - 3. is negative
 - 4. is undetermined



Things to be careful with...

- Avoid double negatives
 - (e.g., I don't like not to eat)
- Special terminology....
 - (e.g., Have you ever been contacted by a GPO, GDQ?;



Things to be careful with...

- A person cannot belong in more than one category (plus, in principle we don't want to categorize continuous variables)

e.g., age

16 - 20 

20 - 25 

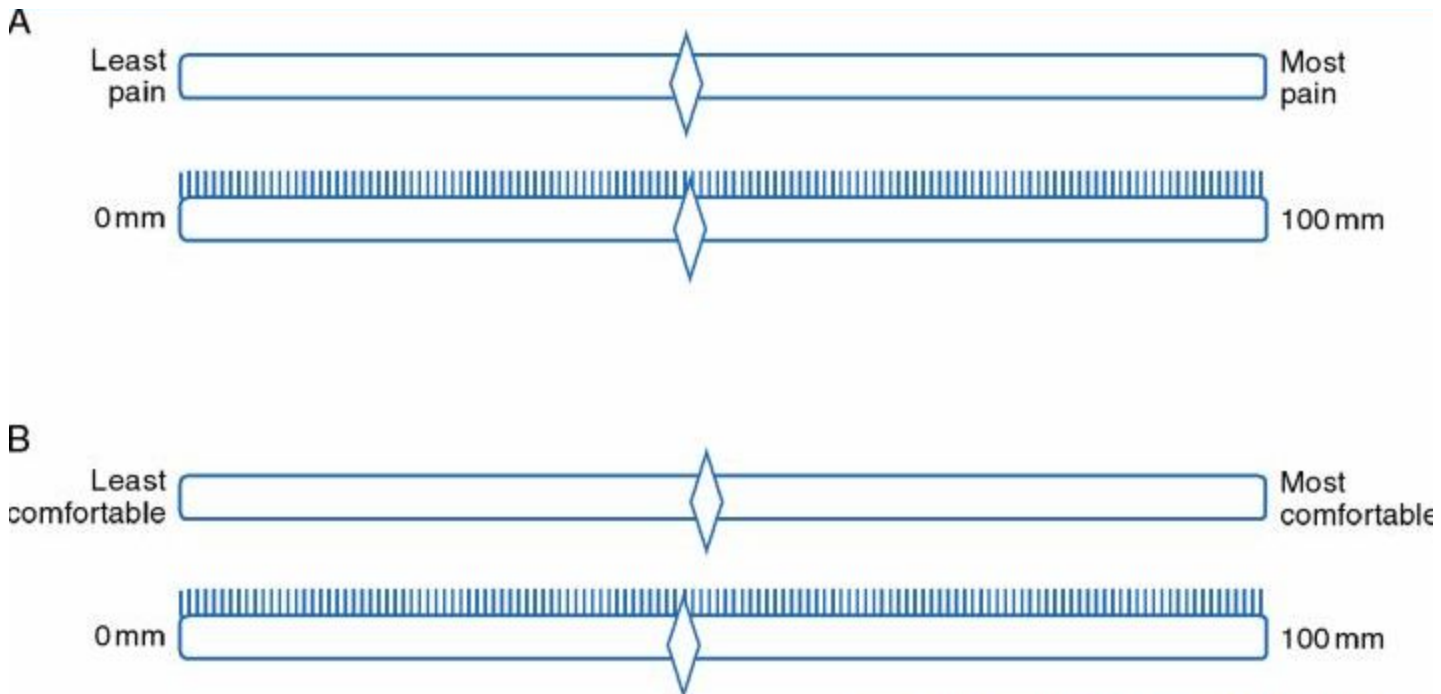
25 - 30 

30 - 35 



Things to be careful with...

- So more numbers is desirable, if they are good numbers



Things to be careful with...

- Avoid questions that “lead” to an answer
e.g.

Do you agree that health care is under-financed?

- Avoid strong words like «always», «never», etc. replace them with «sometimes» or «oftentimes» or *let the response option define the strength of the response.*
- e.g., I strongly believe that health care is under-financed
- e.g., I believe that health care is under-financed
- Strongly Disagree Disagree Agree Strongly Agree



Things to be careful with...

- Avoid emotion eliciting words
 - E.g., A learning disability is a disorder of personality with implications for social functioning
- Check that there is only one correct responding
 - E.g., Alcoholism is a:
 - 1. Disease
 - 2. Habit
 - 3. Disorder of dependency
 - 4. Way to have fun



Things to be careful with...

- **Do not prime the respondent towards a certain response**

NO: How much do you like *or dislike* our **new and improved** reception area?

not at all <----->very much so

YES: How much do you like the reception area?

not at all <----->very much so



Things to be careful with...

- Avoid complex patterns of responding
 - E.g., A and B; A and C; All the above; None of the above
- Unless you want to make things too difficult
- But be prepared to induce measurement error due to complexity (erroneous responses due to complexity and not to lack of ability)



Things to be careful with...

- Right and wrong answers should be of the same length
 - **It has been observed from empirical studies that correct responses are longer, compared to incorrect**



Things to be careful with...

- Use positively worded rather than negatively words items

Wrong:

- E.g. Which of the concepts below *does not relate* with mastery goals
 1. Avoidance goals
 2. self-determination
 3. Other reference
 4. All of the above
 5. None of the above

The negative stem (does not relate) may confuse the respondents

Wright:

- Which of the concepts below is part of the operational definition of mastery goals
 1. interest
 2. efficacy
 3. flow
 4. all of the above
 5. none of the above



Things to be careful with...

- In multiple choice items use 4-6 choices.
- Odd numbers of options may be associated with unweighted schemes (more positive or more negative options), assuming that we do not want a neutral option.
- Avoid “don’t know” “don’t understand” “neither..nor..”
- “Not applicable”.
- Place correct responding in every possible place among items (not all correct on Bs).



Things to be careful with...

1	2	3	4	5
No impact: Did not change your career course or opportunities		Moderate impact: opened some new career avenues or opportunities		Significant impact: altered your career course through change in direction or new opportunity

- Scaling with undefined response options

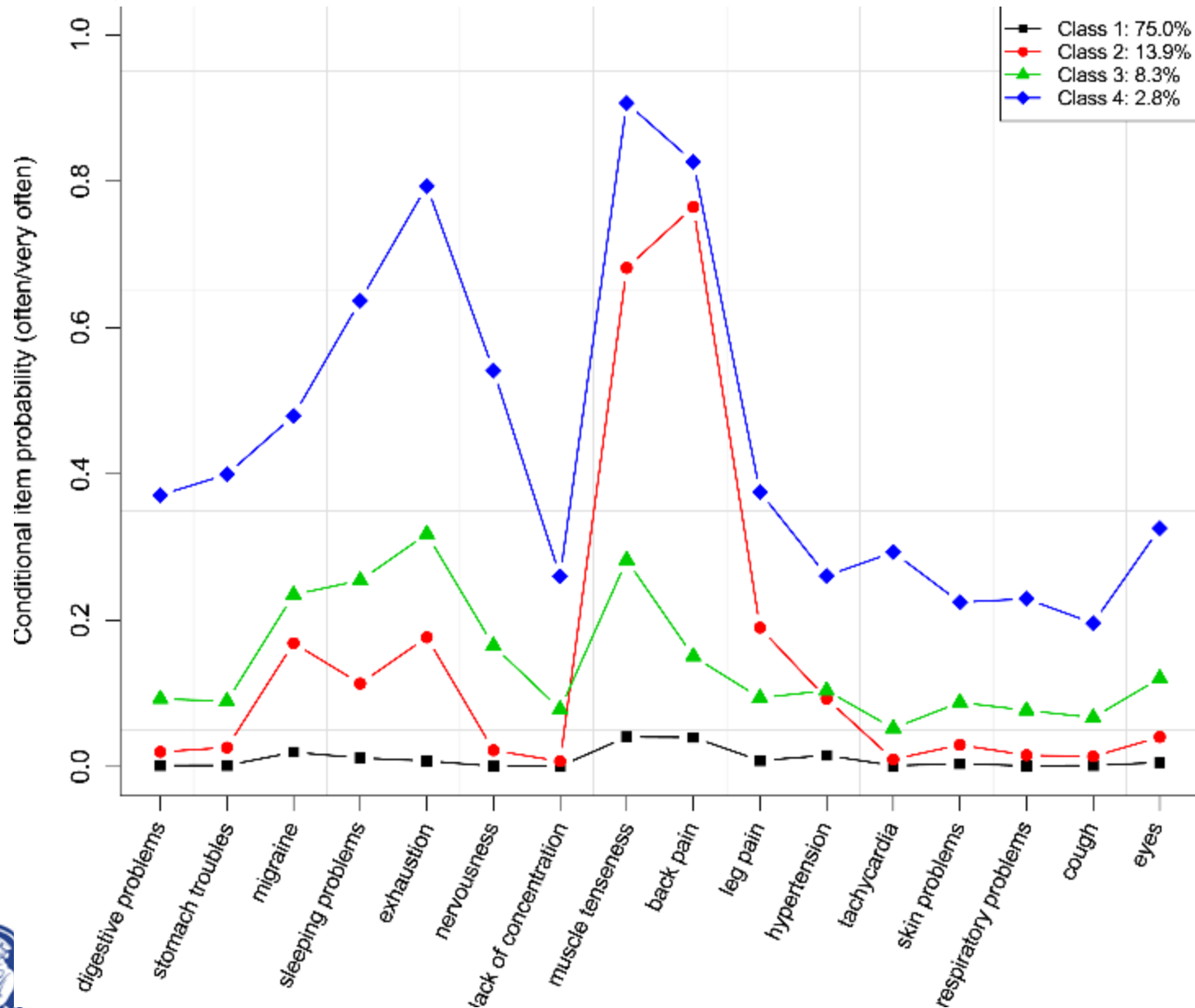


Things to be careful with...

- Please assess the degree of impact this rotation had on your personal growth as a physician.
- Operational definition of personal growth
- Measurement of one-item variables



Things to be careful with...



Things to be careful with...

- Completion Strategy
- Skip items
- Dual Responding

How important are the following issues for you personally?
Circle the number that best represents your opinion on a scale from 1 (extremely unimportant) to 10 (extremely important)

	Extremely unimportant										Extremely important
Protection of endangered species	1	2	3	4	5	6	7	8	9	10	
Improving air quality	1	2	3	4	5	6	7	8	9	10	
Creating renewable energy sources	1	2	3	4	5	6	7	8	9	10	
Improving water quality	1	2	3	4	5	6	7	8	9	10	
Reducing the use of nuclear power	1	2	3	4	5	6	7	8	9	10	



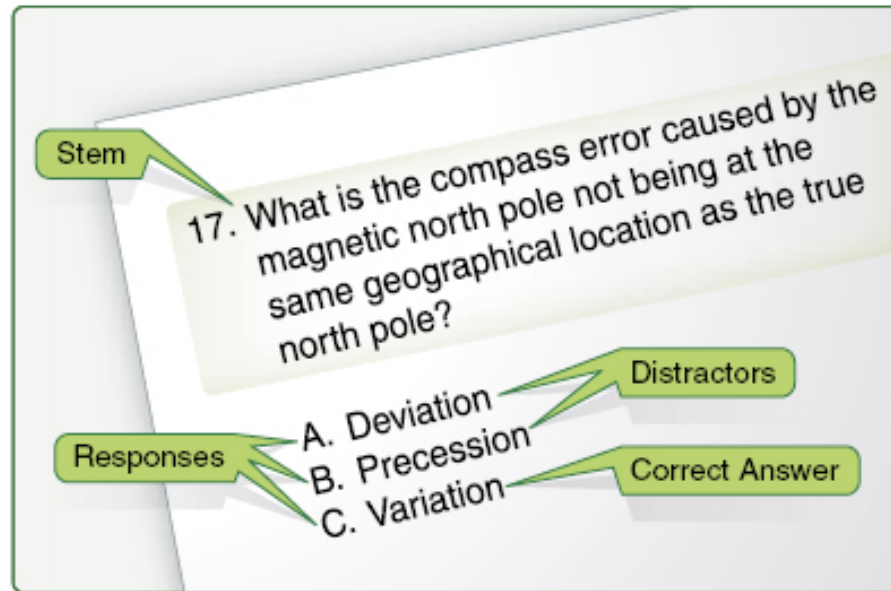
Things to be careful with...

- What is your current marital status? (Check only one)
 - ☐ Married or long term committed relationship
 - ☐ Divorced or separated
 - ☐ Single/never married
 - ☐ Widowed
- Is this important to measure? Part of focal research questions?



Things to be careful with...

- Manipulate item difficulties to avoid floor and ceiling effects
- Distractors and Item Difficulties



Things to be careful with...

- Measurement of depression, items:
- I hate going out
- I hate to socialize with other people
- It is hard for me to make friends
- It is hard for me to not stand out from the crowd
- It is important for me to be happy (R)
- It is important for me to have friends (R)
- I often feel left out
- I often feel rejected
- What does reversing negatively
- Worded items mean?
- e.g., Rosenberg's self esteem scale
- Method variance

	Questions	Strongly agree	agree	disagree	Strongly disagree
1	On the whole I am satisfied with myself.				
2*	At times I think that I am no good at all				
3	I feel that I have a number of good qualities.				
4	I am able to do things as well as most other people				
5*	I feel I do not have much to be proud of				
6*	I certainly feel useless at times.				
7	I feel that I am a person of worth, at least the equal of others.				
8*	I wish I could have more respect for myself.				
9*	All in all, I am inclined to feel that I am a failure.				
10	I take a positive attitude toward myself.				



Presentation and Other Issues

- Socially desirable responding?
- Clear instructions, example on how to complete the scale/respond
- Length of test (fatigue, boredom, loss of interest/motivation, etc.)
- Place of administration
- Seating arrangement
- Lighting
- Time of testing
- If the respondent is not feeling well?
- Stop the test at any time
- Pilot test



Other Issues that potentially contribute to measurement error

- Indirect sources of information:
- Ask principals about teachers
- Ask hospital directors about doctors, etc.
- Always ask the source that is best suited to get the best possible information.
- Complex designs? Assess caregivers: who to ask? Mothers or fathers? Or both? Do they have the same perception and experience? How to analyze data?



Summary

- Theory, how many constructs
- Item development, how many items per construct? Content validity first, develop more items than are needed, item pool
- We try not to categorize continuous variables
- Item scaling, 5-7 response option? Middle response delete it is not in the middle of the continuum between – and + infinity
- Reversely coded items? Good to check participant's attention but usually with measurement error
- Avoid double negatives
- Avoid more than one idea in any one item
- Avoid different scaling options for different items
- Avoid dichotomous options
- Power analysis for pilot and validation study
- Pilot test what? (procedures, item wording, ethics, fatigue, participant experience etc.)
- Design for various forms of validity
- Socially desirable responses
- Ceiling – floor effects; with socially desirable surveys we have ceiling effects, little variance



Our fight is to minimize error...

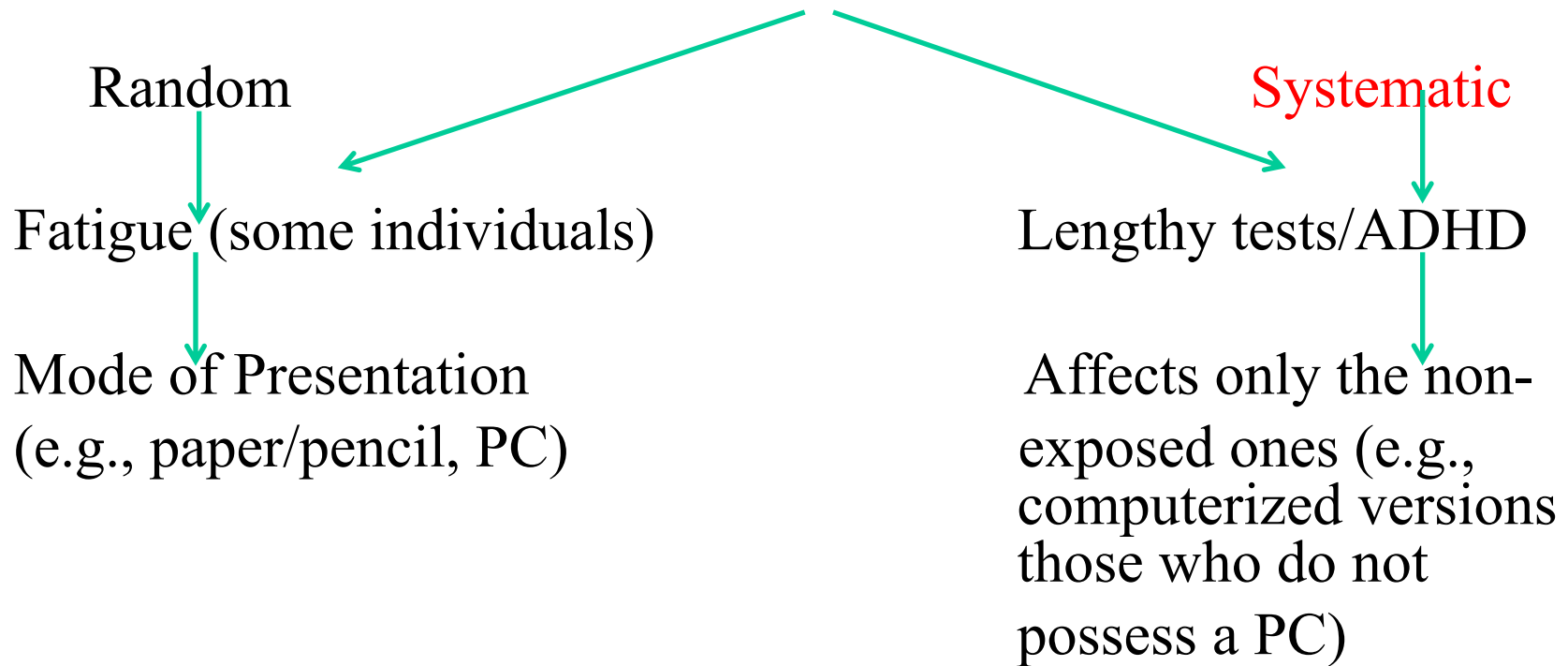


Children's Hospital Boston

Conceptualization of Measurement Error

Measurement Error and Classical Test Theory

$$\text{True} = \text{Obs.} + e$$



Systematic measurement error is due to a third variable and confounds the measurement of the latent trait of interest (e.g., aptitude).



Non-Test Related Factors Associated with Measurement Error

Contextual Effects

1. Different teams employ different criteria in their decision process
2. Bias related to groups (schools, communities)

Individual Effects:

1. Bad day, bad mood
2. Fatigue, illness
3. Motivation (can be good or bad)
4. Stress (can be good or bad)
5. Setting (familiar and/or promoting or demoting performance)



Test Related Factors Associated with Measurement Error

1. Directions (confusing or additional clarifications)
2. Cultural (a correct response is considered incorrect in one culture compared to the other)
3. Religious (content having different connotations in different religions)
4. Test Length (related to concentration or fatigue)
5. Test Content (e.g., offensive)
6. Test Format (familiar vs. unfamiliar)
7. Unidimensionality (e.g., math problem with verbal directions)



Reliability



Children's Hospital Boston

Reliability and Validity in Measurement



‘Validity presupposes reliability’

An instrument can be reliable by not valid

An instrument cannot be valid without being reliable



Evaluating the Quality of a Measure - Reliability

Repeated measurements of the same object with the same measure are the same.

→ Stability and consistency of the measure

- Agreement of measure with itself on different occasions / Consistency of measure (Test-Retest Reliability)
- Stability of responses across measures in the same test (Internal Consistency)



Reliability: Test-Retest Reliability

- Assumptions:
 - No actual change occurred
 - Second measurement (M2) is not influenced by first measurement (M1)
- Time interval between measures should be carefully chosen!



Reliability: Test-Retest Reliability

- How to Evaluate:

Pre	Post	Agreement
5	6	No
4	5	No
12	13	No
7	7	Yes
5	5	Yes

Reliability: $\text{Agreements} / (\text{Agreements} + \text{Disagreements})$

$$= 2 / (2 + 3) = 2 / 5 = 40\% \text{ reliability}$$

- Is this too harsh? Do you want to use a range of scores that defines consistency?
- How much is the estimate of reliability if you define a range of ± 1 between pre-posttest as still being consistent?



Reliability: Internal Consistency

- General procedure:
 - Assess correlation between different measures (items) within a test/survey/questionnaire.
 - Only one survey administration needed.
 - Assumption: Items are supposed to measure the same dimension of the construct the researcher wants to measure.
- Measures:
 - Cronbach's Alpha $\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$
 - Kuder-Richardson 21 formula for dichotomous items



Internal Consistency Versus Content Validity

Measurement of AGQ (Elliot & McGregor, 2001)

Table 1
Study 1: Factor Loadings for Achievement Goals

Achievement goal item	Factor			
	Performance approach	Mastery avoidance	Mastery approach	Performance avoidance
1. It is important for me to do better than other students.	.93 (.97)			
2. It is important for me to do well compared to others in this class.	.89 (.90)			
3. My goal in this class is to get a better grade than most of the other students.	.89 (.91)			
4. I worry that I may not learn all that I possibly could in this class.		.90 (.93)		
5. Sometimes I'm afraid that I may not understand the content of this class as thoroughly as I'd like.		.86 (.88)		
6. I am often concerned that I may not learn all that there is to learn in this class.		.84 (.85)		
7. I want to learn as much as possible from this class.			.91 (.93)	
8. It is important for me to understand the content of this course as thoroughly as possible.			.90 (.93)	
9. I desire to completely master the material presented in this class.			.80 (.78)	
10. I just want to avoid doing poorly in this class.				.87 (.90)
11. My goal in this class is to avoid performing poorly.				.85 (.88)
12. My fear of performing poorly in this class is often what motivates me.				.74 (.70)

Assumptions of Cronbach's alpha (a) all questions are indicators of the same construct, and, (b) the content of the items is independent (i.e., the answer to one item does not influence how the respondent responds to another).



Reliability: Concepts

ID	Pre	Post
1	1	3
2	6	8
3	12	14
4	15	17
5	4	6
6	7	9

- Reliability = Stability
- Types:
- Test-retest, kappa, correlation. Time lag between measurements?
- Correlation as a mean to assess reliability
- Alternate forms of tests (two errors due to time and due to test content)
- Split-half, odd/even, less reliable due to test length
- Both of the above assess content sampling and individual differences in behavior (heterogeneity of behavior)
- Reliability between raters?
- Reliability between observers?



Omega Reliability Using Online Calculator

- <http://www.thestatisticalmind.com/calculators/comprel/comprel.htm>

Composite Reliability Calculator

Estimates Composite Reliability based on Standardized Factor Loadings and Error Variances

Item Number	Standardized Loading	Error Variance	Item R-Square	
1	.96	0.078	0.922	
2	.923	0.148	0.852	Delete
3	.752	0.434	0.566	Delete
4	.747	0.442	0.558	Delete
5	.707	0.500	0.500	Delete
6	.663	0.560	0.440	Delete

Add Item

Composite Reliability: 0.913

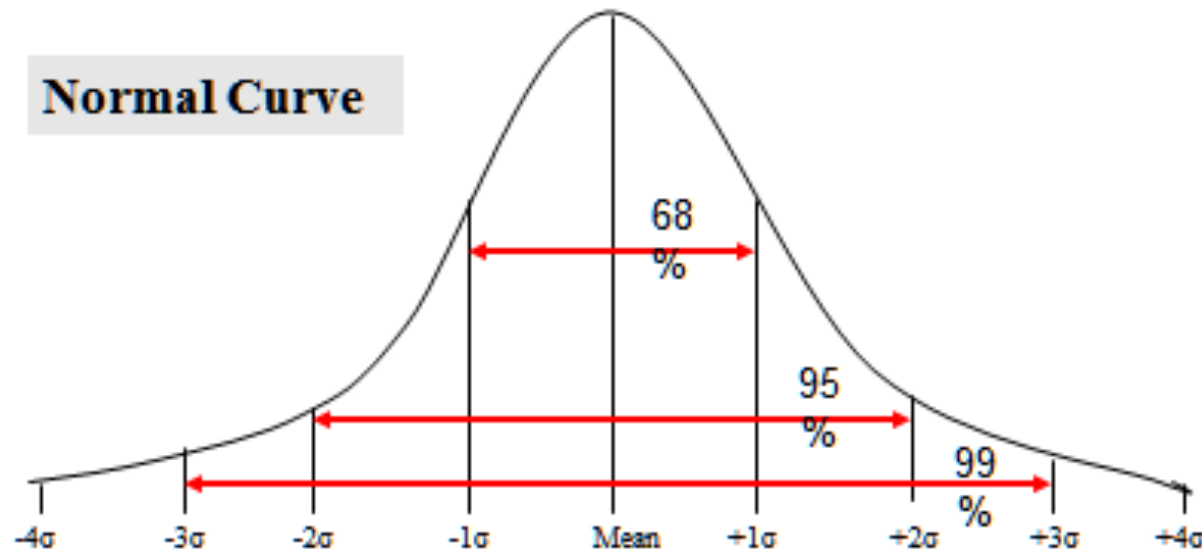
Reset Form



Standard Error of Measurement

- Let's assume that the score in IQ of a person is 100 units (mean)
- It's role in measuring imperfect measures (+-)
- $$sem = SD\sqrt{1 - r_{tt}} = 15\sqrt{1 - .89} = 15\sqrt{.11} = 15 * .33 = 4.97$$
- For $\alpha = .89$ and $sd = 15$, $sem = 5$
- ± 5 is the score of a person at confidence 68%
- ± 10 is the score of a person at confidence 95%
- ± 15 for 3 sds is the score of a person at confidence 99%.

Normal Curve



Factors that Affect Reliability

- Actual change of behavior
- Changes due to familiarity with instrument
- Changes due to large interval
- Changes due to memory of measured construct
- Changes due to first measurement acting as practice
- Changes due to real practice - individuals trying to mess up your study
- Fatigue
- Number of items (bias in Cronbach's alpha)
- Multidimensionality – confusing items (math problem)
- Small or large response option scheme (unknown effect on r)
- Outliers when using correlations



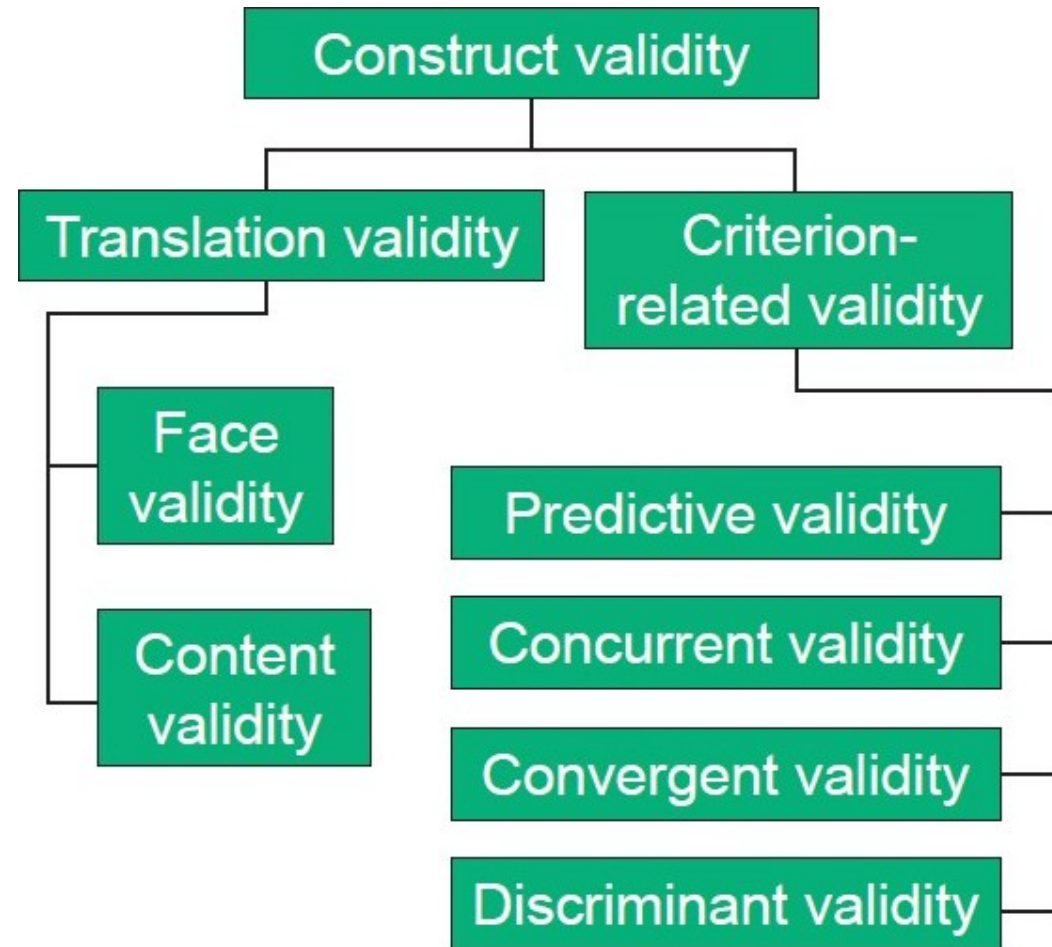
Validity



Children's Hospital Boston

Validity Types

- “Construct” validity
- “Content” validity
- “Convergent” validity
- “Statistical conclusion” validity
- “Incremental”
- “Concurrent” validity
- “Predictive” validity
- “Criterion-related” validity
- “Discriminant” validity



Characteristics of Reliability and Validity

- Has to be applied to all measurement
- It is not a property of the test but the sample!
- It is not a one time thing...how often should we evaluate it?



Content

- (Content validity)
 - All available constructs-theory
- In **content validity**, you essentially check the operationalization against the relevant content domain for the construct. This approach assumes that you have a good detailed description of the content domain, something that's not always true.
- For example, in the assessment of numerical skills at the elementary education level you cannot leave out division or multiplication.



Criterion-Related Validity

- In **criteria-related validity**, you check the performance of your operationalization against some criterion. We usually make a prediction about how the operationalization will *perform* based on our theory of the construct.
 - Test related to a criterion (can be another test or a behavior)
 - Could take place at same time (Concurrent validity)
 - Or in the future (Predictive validity)

In **predictive validity**, we assess the instrument's *ability to predict something it should theoretically be able to predict*. For instance, we might theorize that a measure of math ability should be able to predict how well a person will do in an engineering-based profession. We could give our measure to experienced engineers and see if there is a high correlation between scores on the measure and their salaries as engineers.



Discriminant Validity

- In **discriminant validity**, we examine the degree to which the operationalization is different for different populations. That is we assess discriminant validity by the ability of the measure to differentiate different groups of known characteristics (e.g., an IQ measure for kids with ASD vs. typical)



Convergent/Divergent

- (Convergent validity)
 - Positive relationship with similar measurements
- (Divergent validity)
 - Should not correlate with other non theoretically related measurements

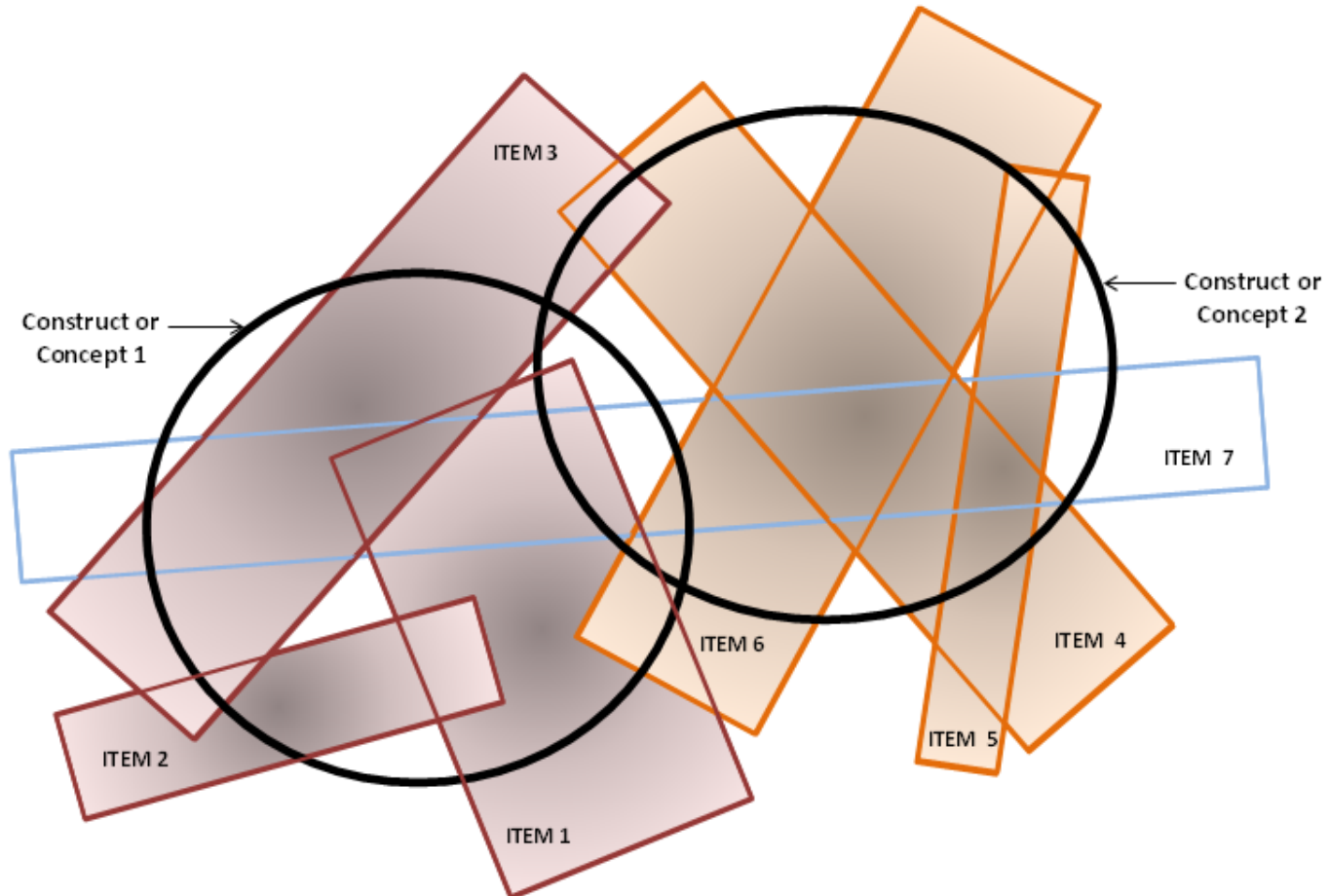
**How strong should the correlation be? (.30?) (.50?) (.75?)

**How weak should the correlation be? .30 just for shared method variance



Construct Validity

- Evaluates relationships between constructs
- Usually evaluated using factor analytic procedures



Evaluating the Quality of a Measure - Validity

- Construct Validity
 - Principal Component Analysis, Common Factor Analysis, Analysis of Covariance Structures
 - Multitrait-Multimethod Assessment



Positive and Negative Affect Schedule - Exploratory Factor Analysis....the old days

Table 2 Item Composition of the PANAS-X Scales

General Dimension Scales

Negative Affect (10)

Positive Affect (10)

afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset, distressed
active, alert, attentive, determined, enthusiastic, excited, inspired, interested,
proud, strong

Basic Negative Emotion Scales

Fear (6)

Hostility (6)

Guilt (6)

Sadness (5)

afraid, scared, frightened, nervous, jittery, shaky
angry, hostile, irritable, scornful, disgusted, loathing
guilty, ashamed, blameworthy, angry at self, disgusted with self, dissatisfied with self
sad, blue, downhearted, alone, lonely

Basic Positive Emotion Scales

Joviality (8)

Self-Assurance (6)

Attentiveness (4)

happy, joyful, delighted, cheerful, excited, enthusiastic, lively, energetic
proud, strong, confident, bold, daring, fearless
alert, attentive, concentrating, determined

Other Affective States

Shyness (4)

Fatigue (4)

Serenity (3)

Surprise (3)

shy, bashful, sheepish, timid
sleepy, tired, sluggish, drowsy
calm, relaxed, at ease
amazed, surprised, astonished

Note. The number of terms comprising each scale is shown in parentheses.

Table 11 Varimax-Rotated Factor Loadings of the Positive Mood Terms Defining the Joviality, Self-Assurance, Attentiveness, and Serenity Factors (Past Week Instructions, N = 607)

Descriptor	Factor			
	1	2	3	4
happy	.79			
cheerful	.78			
joyful	.78			
excited	.77			
enthusiastic	.73			
lively	.70	.30		
energetic	.64		.31	
delighted	.63			
bold		.71		
fearless		.69		
strong		.68		
proud		.64		
confident		.60		.34
daring	.42	.57		
concentrating			.79	
attentive			.77	
determined			.66	
alert			.61	
calm				.78
relaxed				.76
at ease	.37			.62

Note. Loadings below .30 are omitted.



Children's Hospital Boston

A Confirmatory Factor Analysis Model

