

Constructing Measures to Assess Proficiency in Less Commonly Taught Languages (LCTLs)

What Bikes, Shirts, and Hurdles Teach Us About Assessment

Troy Cox, Ph.D.
Associate Director of Research and Assessment,
Center for Language Studies
Brigham Young University

Introductions

- In the CHAT box, please type your
 - Name
 - Affiliation
 - Language(s)

Troy L. Cox, Brigham Young University, TESOL & Italian (Chinese, Russian, Spanish, French)

Disciple•Husband•Father•Amateur Cyclist, Yogi, Cook & Photographer•Professor (Applied Linguistics/Language Testing)



Poll 1

- How comfortable are you with general principles of assessment including the statistics used in the field?
 - Very comfortable
 - Somewhat comfortable
 - Neutral
 - Somewhat uncomfortable
 - Very uncomfortable

Poll 2

- How comfortable are you with language assessment including the four skill areas as well as the subskills?
 - Very comfortable
 - Somewhat comfortable
 - Neutral
 - Somewhat uncomfortable
 - Very uncomfortable

Poll 3

- How comfortable are you with describing the ACTFL proficiency guidelines to colleagues and students?
 - Very comfortable
 - Somewhat comfortable
 - Neutral
 - Somewhat uncomfortable
 - Very uncomfortable

An instrument is always something **secondary**:
There is always a **purpose** for which an instrument is
needed and a **context** in which it is going to be used.

Emphasis Added (Wilson, 2005, p. 6)

What is your
context?

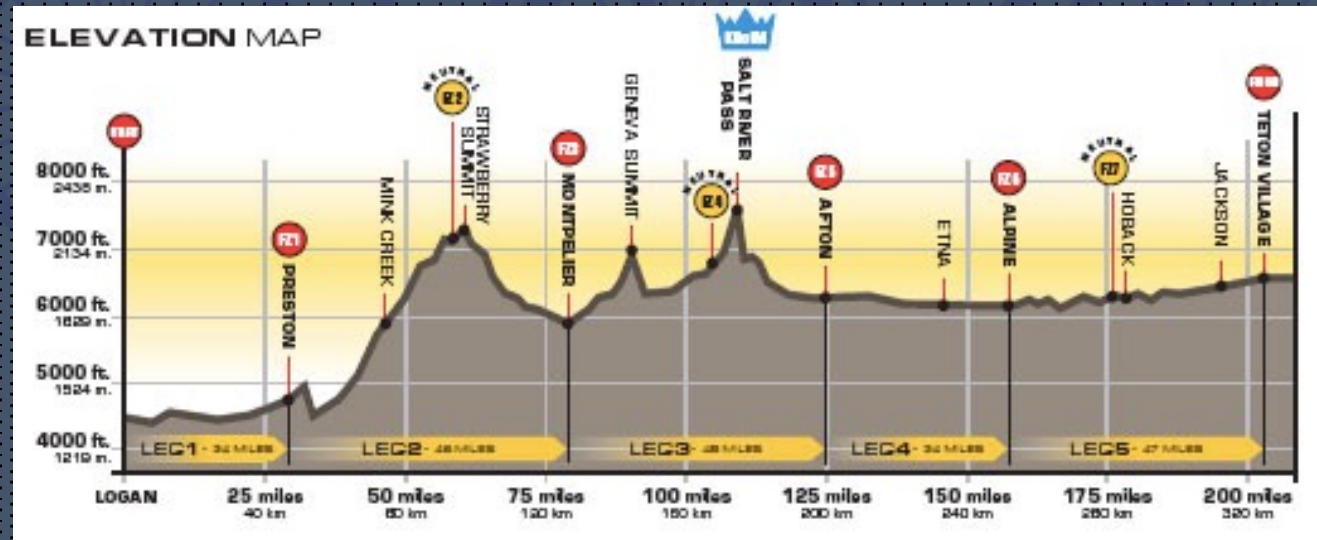
What is
your
purpose?

Criterion-Referenced vs. Norm-Referenced Tests

Bicycles

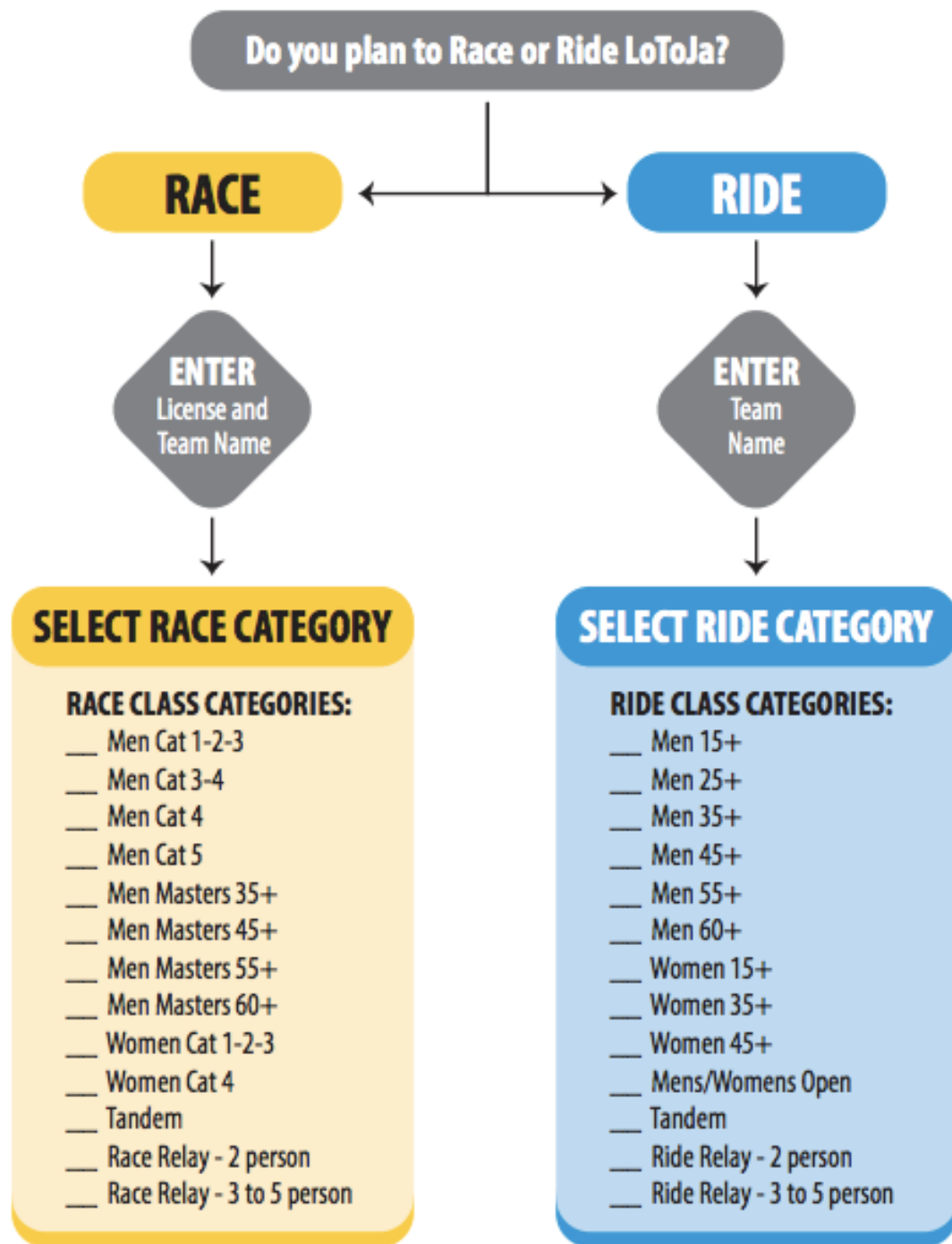


- The LoToJa Bicycle Classic is a 206-mile (332 km), one-day amateur bicycle road race from **Logan**, UT **to** **Jackson Hole**, WY, USA.



Is LOTOJA a criterion or norm-referenced competition?

Norm-
referenced



Criterion-
referenced



If you're competing against others, then



LOTOJA 2014 RIDE SPEEDS AND TIMES - NOMINAL									
Stage	Total Stage Miles	Flat Miles	Climb Miles	Steep Climb Miles	Descend Miles	Stage Time	Arrive Time		
Logan to Preston (34)	34	33	1			1.60	7:48 AM		
Preston to Montpelier (80)	46	6	20	6	14	3.13	11:08 AM		
Montpelier to Afton (125)	45	7	14	6	18	2.89	2:13 PM		
Afton to Alpine (158)	33	28	5			1.65	4:04 PM		
Alpine to Finish (205)	47	37	10			2.41	6:42 PM		
TOTAL Terrain Miles	205	111	50	12	32	11.7			
Average Speed on Terrain		21.5	14.5	6.0	30.0				
Time on Terrain (hrs)		5.2	3.4	2.0	1.1	11.7			
Total Ride Time on Course	11.7	hrs							
Average Speed on Course	17.6	mph							
Total Elapsed Time w/Stops	12.5	hrs	See stop times below						
Start	6:12 AM								
End	6:42 PM								

Stop	Miles	Type	Food	Time	Activity
start	0		na		Start
1	34	Support	Drink/eat some, load for climbs	6	Drink shake, Leave some warm clothing
2	61	Neutral	Snacks from neutral support	6	Regroup from climb, get water, food in bento
3	80	Neutral	Snacks from neutral support	6	Get water, eat snacks, food in bento
4	106	Neutral	Snacks from neutral support	6	Regroup from climb, get water, food in bento
5	125	Support	Drink/eat lots, snacks for WY	12	Drink shake, eat, cool down
6	158	Support	Drink/eat all possible, load snacks	8	Drink shake, cool down, bring on warmer clothing if needed
7	180	Neutral	Snacks from neutral support	5	Mix drinks, snack if possible
end	206			49	minutes 0.82 hours

If you're competing against the clock (external criteria), then it's criterion-referenced.

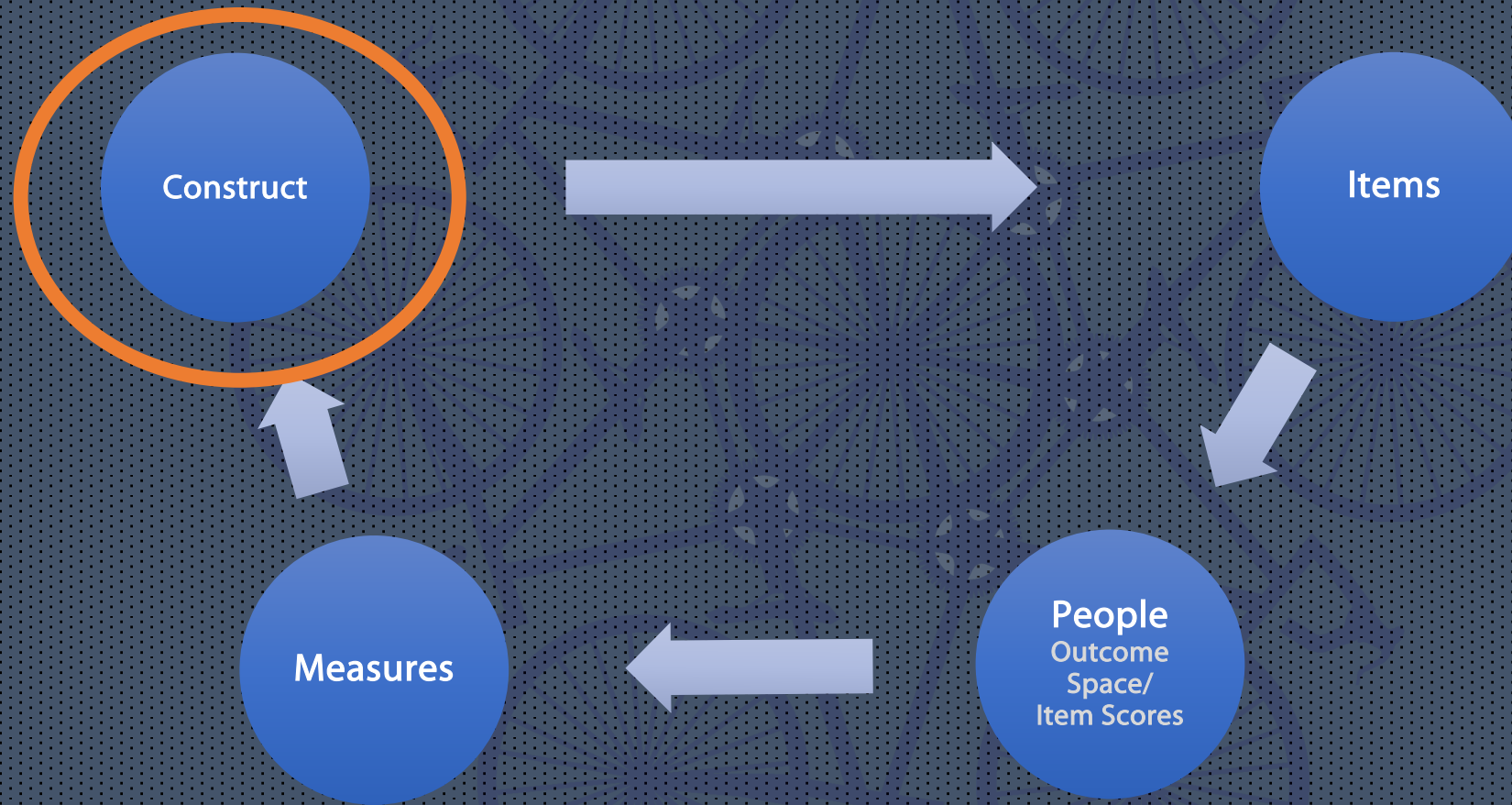


Do you want to measure the LCTL students against each other or to external standards?

Instrument Development Cycle



Instrument Development Cycle



Construct Map

People

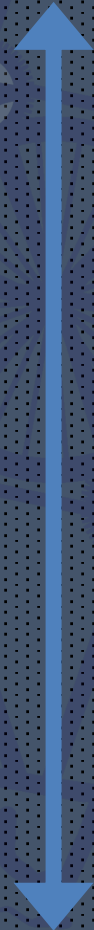
- Do people vary in their ability so that on a continuum, **someone** that has **MORE** ability is likely to be able to do everything than a person with **LESS** ability?

Items/Tasks

- Do tasks vary in their difficulty so that on a continuum, **a task** that is **MORE** difficult is likely to require mastery of everything that is **LESS** difficult?

Not all Traits can be Measured Linearly!

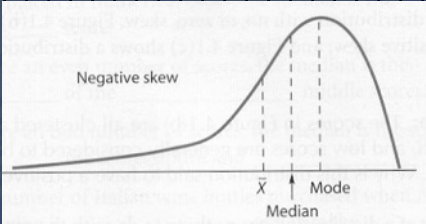
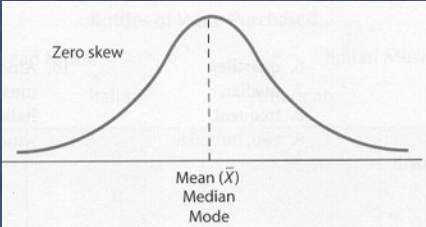
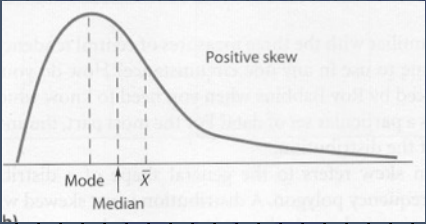
- If you can't define your trait unidimensionally, than you shouldn't claim to measure it.
- Otherwise, what do the numbers you assign mean?



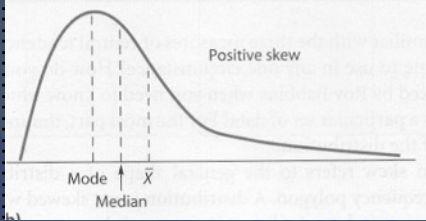
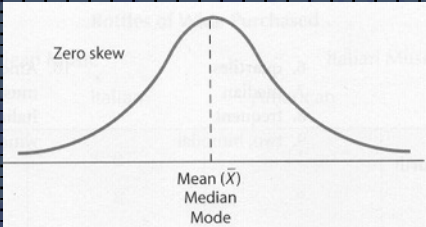
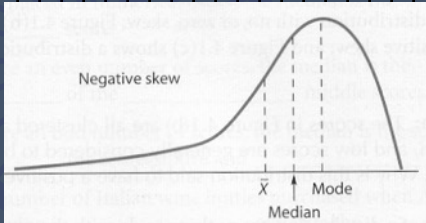
Construct Map

	People	Items/Tasks
	Think of people you know with three different trait levels. How they are different than each other?	Think of tasks with three different levels of ability. How they are different than each other?
		
		
		

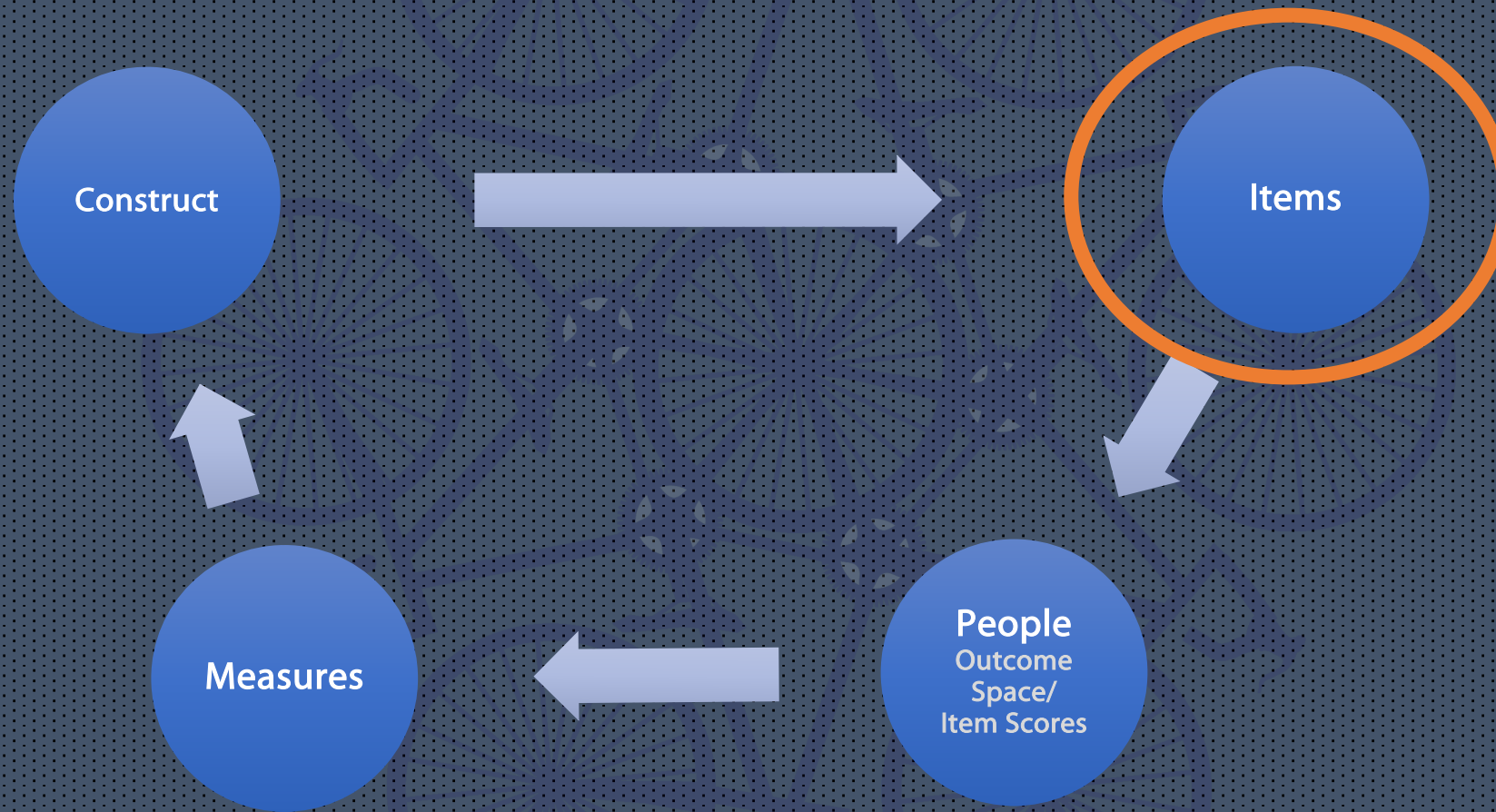
If items/tasks are all at the same level, what would be the distribution of different levels of respondents to those items?

	People	Items/Tasks
	 Negative skew $\bar{X}$ Median Mode	
	 Zero skew Mean ($\bar{X}$) Median Mode	
	 Positive skew Mode $\bar{X}$ Median	

If respondents are all at the same level, what would be the score distribution on different levels of items?

	People	Items/Tasks
		 Positive skew Mode $\bar{X}$ Median
		 Zero skew Mean ($\bar{X}$) Median Mode
		 Negative skew $\bar{X}$ Mode Median

Instrument Development Cycle



Stages/Levels of Pre-specification Needed in Item Development

Qualitative



Quantitative

Item Format	Intent to Measure Construct "X"	Description of item components		Item Specification Level of Detail		
		General	Specific	No Score Guide	Score Guide	Responses
Observation (informal)	Before or After	After	After	After	After	After
General Trends	Before	Before	After	After	After	After
Specific Trends	Before	Before	Before	After	After	After
Open-ended	Before	Before	Before	Before	After	After
Open-ended plus scoring guide	Before	Before	Before	Before	Before	After
Fixed Response	Before	Before	Before	Before	Before	Before

Every instrument should go through a set of developmental stages moves from the general to the specific. Otherwise, developers run the risk of making arbitrary decisions of design components and there is no defense against the criticism that a fixed response format has distorted the measurement.

Unidimensionality

Shirts

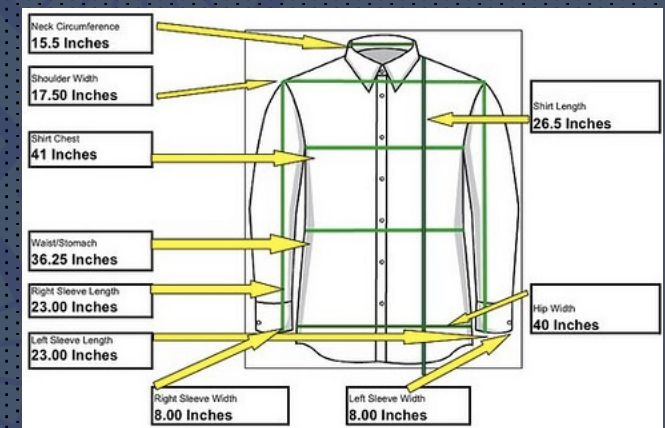
Constructing Measures

- How can you measure something you haven't described?
- Men's Shirt Example
 - How would you measure a man's shirt if you didn't have a tape measure?
 - What would be important to describe?



Constructing Measures

- How can you measure something you haven't described?
- Men's Shirt Example
 - How would you measure a man's shirt if you didn't have a tape measure?
 - What would be important to describe?



Is this a uni-, bi- or multi-dimensional measurement?



Unidimen-
sional

Bidimensional
Off-the-rack Shirt

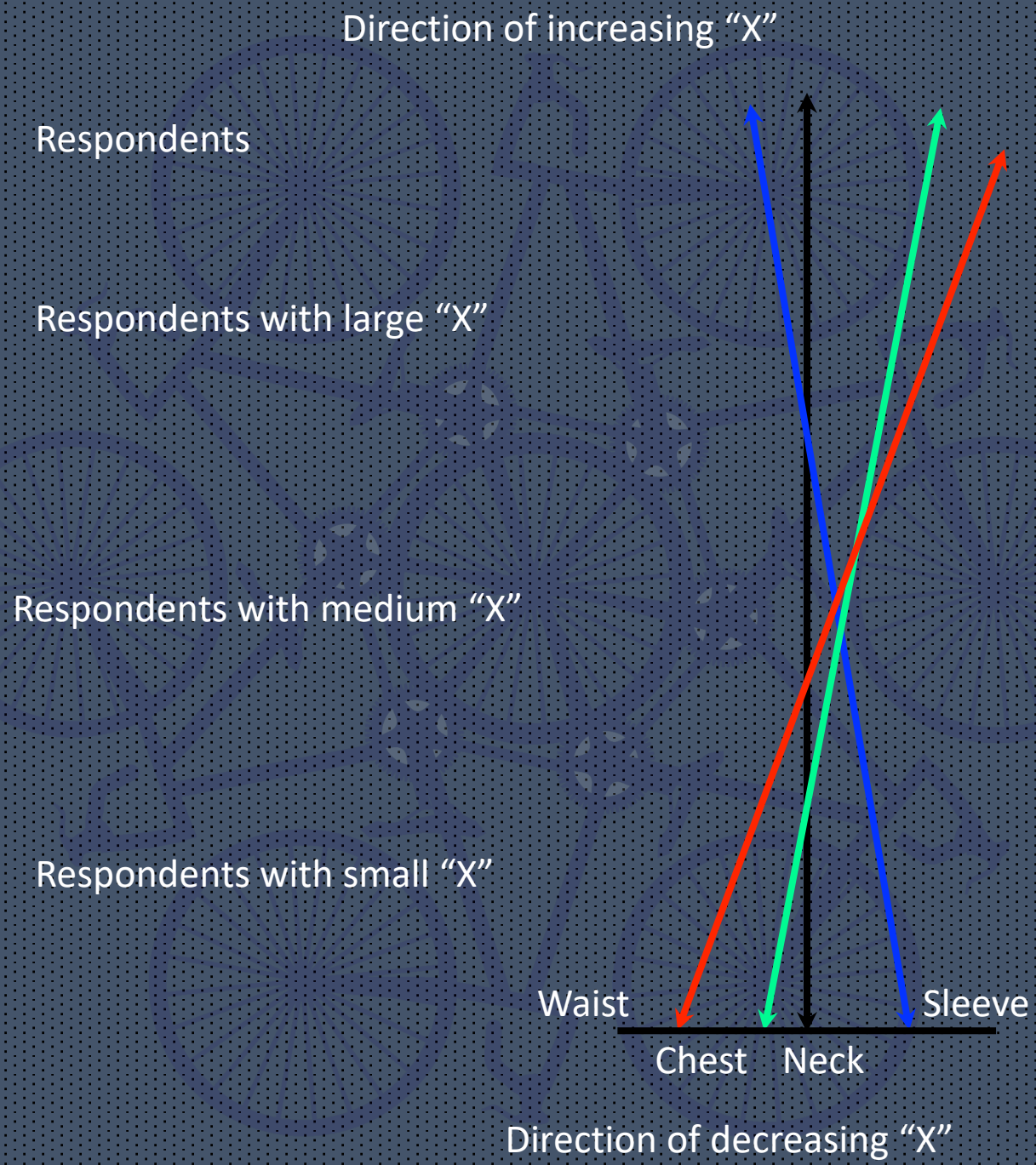
Multidimensional
or the Bespoke,
“Tailored” Shirt

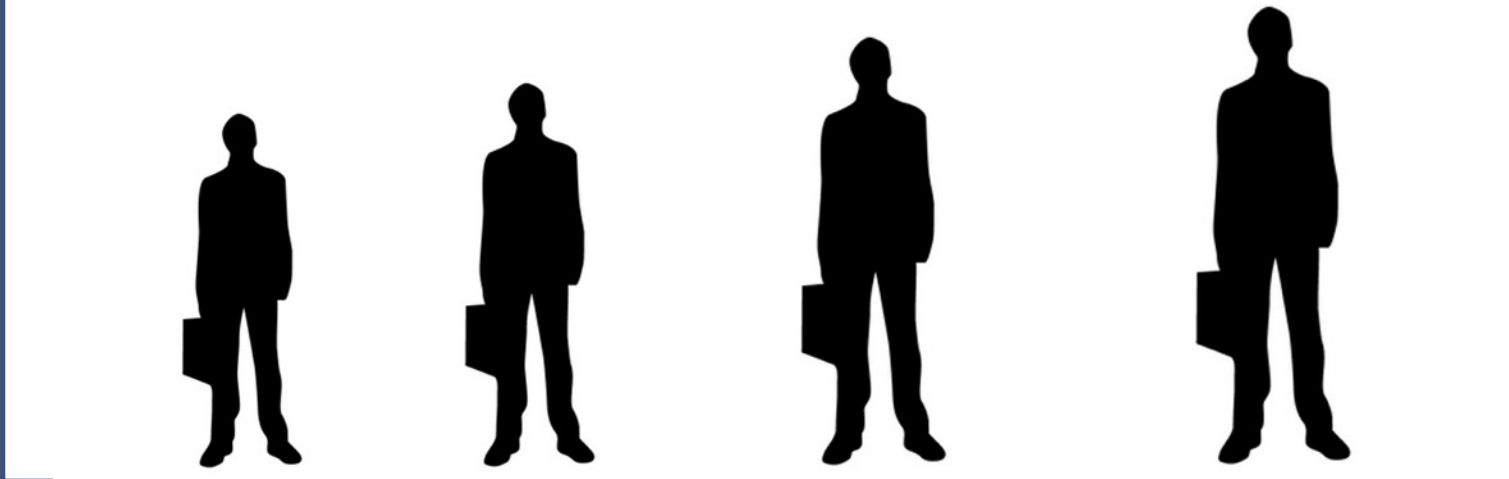
A Venn diagram with three overlapping circles. An orange circle on the left is labeled 'Unidimensional'. A purple circle in the center is labeled 'Bidimensional Off-the-rack Shirt'. A blue circle on the right is labeled 'Multidimensional or the Bespoke, "Tailored" Shirt'. The intersection of all three circles contains a table of shirt sizes.

US Sizes (inches)	Neck	Sleeve	Waist	Chest
S	14-14 ½	32 ½-33	29-31	35-37
M	15-15 ½	33 ½-34	32-34	38-40
L	16-16 ½	34 ½-35	36-38	42-44
XL	17-17 ½	35 ½-36	40-42	46-48

It depends on the
shared
understanding
between
the shirt
manufacturer and
the customer.

Construct Map





Neck, waist, sleeve and chest *tend* to co-occur.

US Sizes (inches)	Neck	Sleeve	Waist	Chest
S	14-14 ½	32 ½-33	29-31	35-37
M	15-15 ½	33 ½-34	32-34	38-40
L	16-16 ½	34 ½-35	36-38	42-44
XL	17-17 ½	35 ½-36	40-42	46-48

What happens when items or people don't fit with the model?



Make informed, deliberate decisions on what you'll do.



US Sizes (inches)	Neck	Sleeve	Waist	Chest
S	14-14 ½	32 ½-33	29-31	35-37
M	15-15 ½	33 ½-34	32-34	38-40
L	16-16 ½	34 ½-35	36-38	42-44
XL	17-17 ½	35 ½-36	40-42	46-48

Compensatory Measurement =

Add scores together and divide to
get the average.

L



US Sizes (inches)	Neck	Sleeve	Waist	Chest
S	14-14 ½	32 ½-33	29-31	35-37
M	15-15 ½	33 ½-34	32-34	38-40
L	16-16 ½	34 ½-35	36-38	42-44
XL	17-17 ½	35 ½-36	40-42	46-48

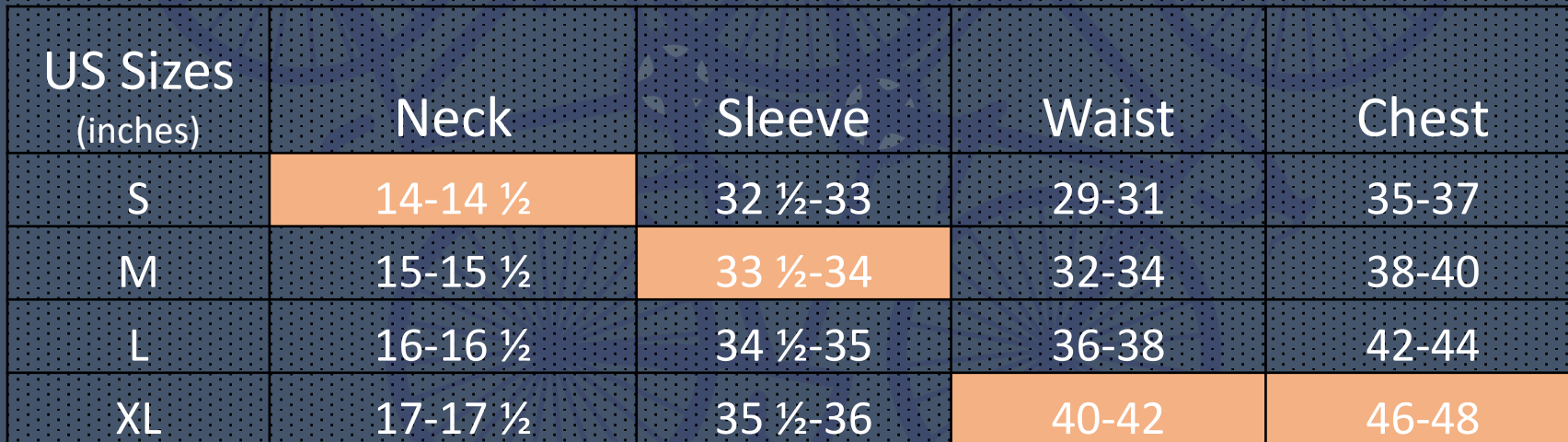
Problem with Compensatory Measurement

	Bilbo	Gandalf	Orin
Neck	14 (S)	15 (M)	17 (XL)
Arm Length	32 (S)	33 (M)	36 (XL)
Chest	42 (L)	40 (M)	35 (S)
Total	88	88	88
	S	M	S

Same score, but three very different profiles.
We do math without thinking about the assumptions.

Measure each component separately and use the smallest

S or XL



Given that consumers will have measurement variation in neck, sleeve, waist and chest size, successful manufactures will ensure that all of their products construct their shirts based on their stated standards.

US Sizes (inches)	Neck	Sleeve	Waist	Chest
S ↔	14-14 ½	32 ½-33	29-31	35-37
M ↔	15-15 ½	33 ½-34	32-34	38-40
L ↔	16-16 ½	34 ½-35	36-38	42-44
XL ↔	17-17 ½	35 ½-36	40-42	46-48

Sizes Are Relative to Population



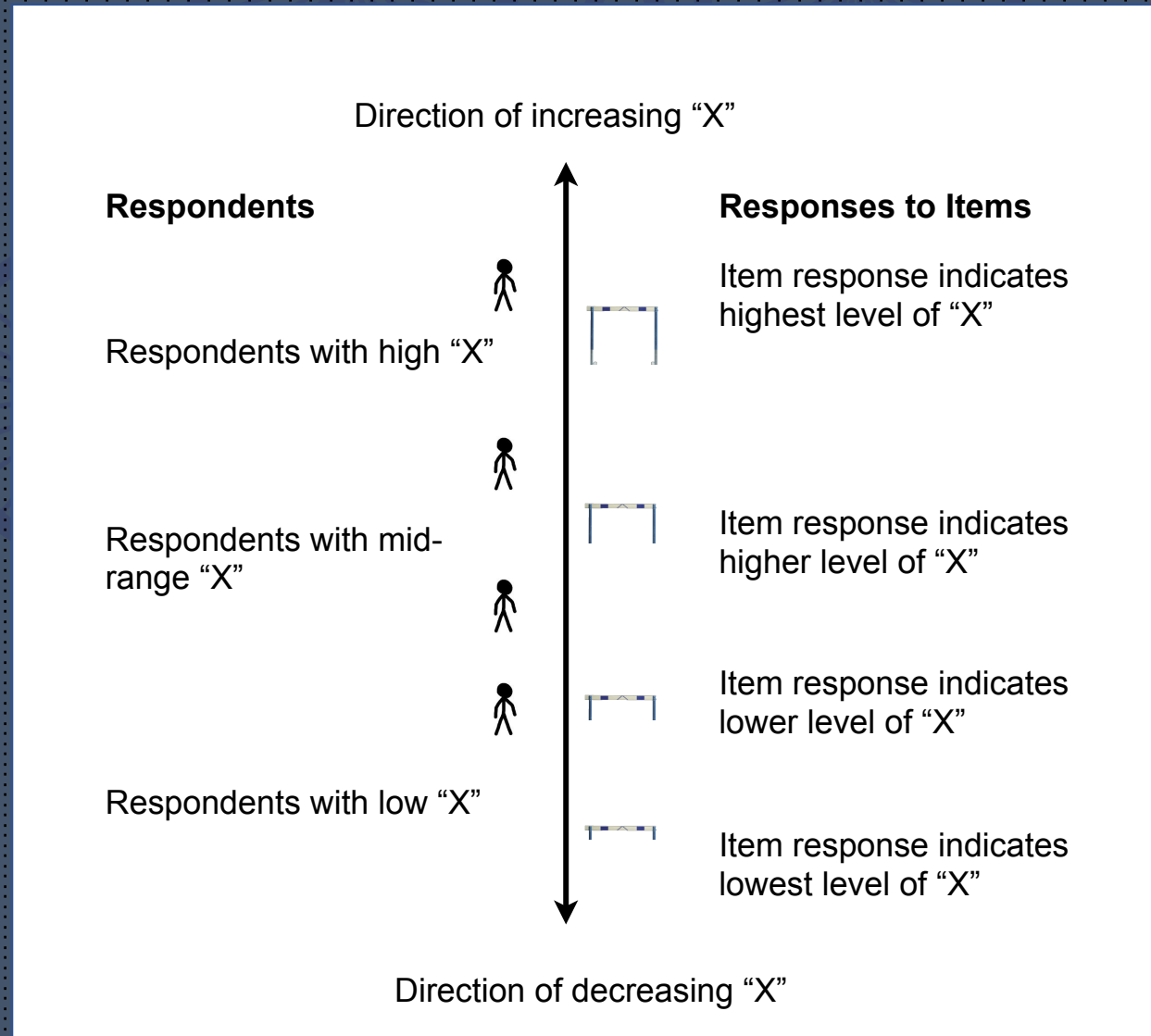
Unidimensionality

- Construct
 - Make sure you (e.g. the manufacturer) and your end-users (e.g. the customer) have a shared definition of your construct map.
- What's XXXL in one context may simply be L in another.

Local Independence

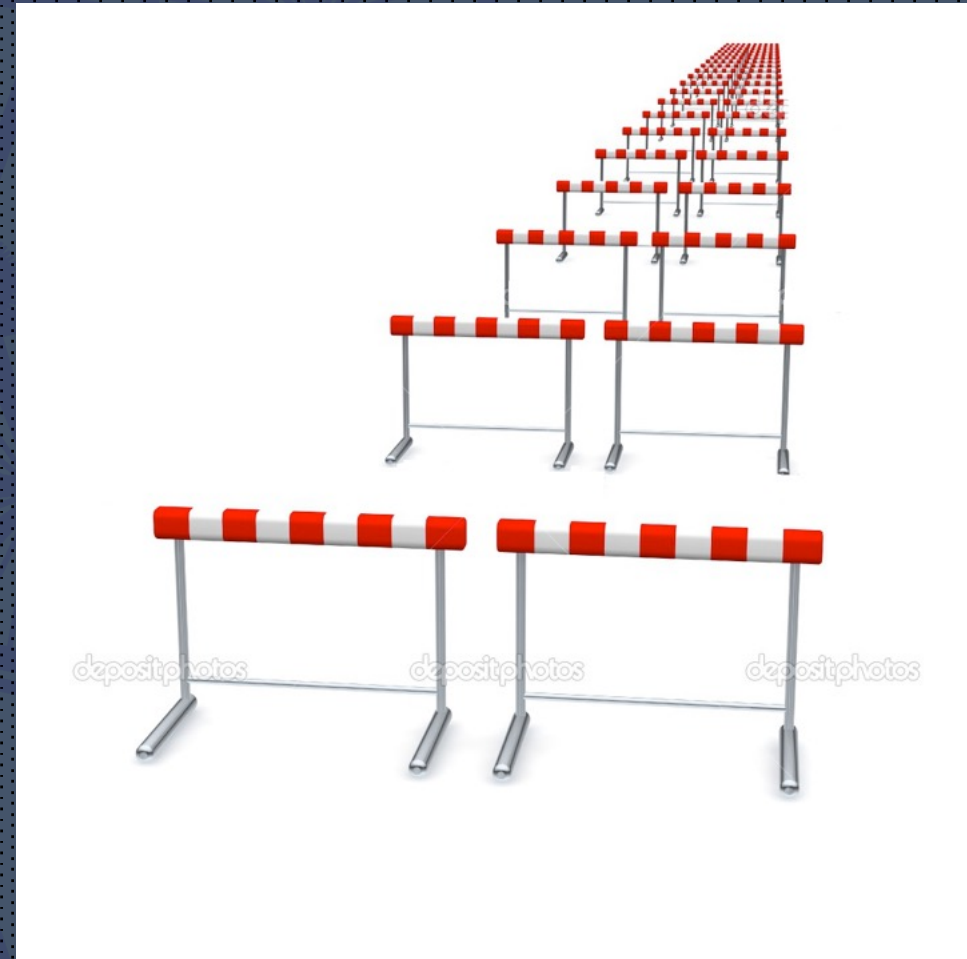
Hurdles

Construct Map Example2 —Jumping



Local Independence

- Repeated performance increases confidence in reliability
- Independence necessary so we **add** the results into a single score



The minute we assign a number to something, we assume we can do math without verifying the assumptions.

For norm-referenced tests, have a range of obstacles that can differentiate the jumpers.



For criterion-referenced tests, have each hurdle align with the desired criteria.



Classical Test Theory

Mathematical Model

True Score = Observed Ability (on entire test)

+

Error (single value for test and test-takers)



Classical Test Theory Limitations

Item Dependent (Person Score is additive result of performance on all items which contribute equally to the score)

Group Dependent (Item Difficulty is proportional result of population of test-takers)

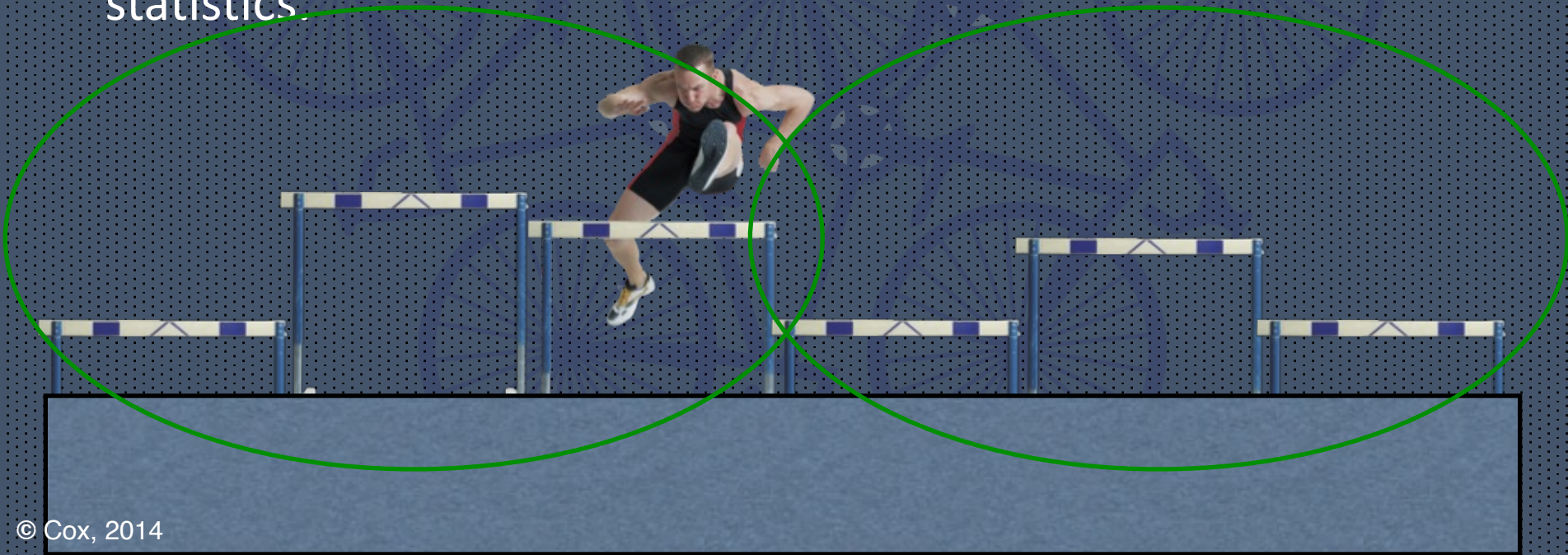


Classical Test Theory Limitations

Only applies to the test being administered.

EACH item is counted as a unit of measure (or interval) on the scale

For test forms to be equated, there need to be shared items. Test forms cannot be equated with just item statistics



PROBLEM: There is NO external norm to validate the measurement instrument

Ideal Interval Level Test "Ruler"



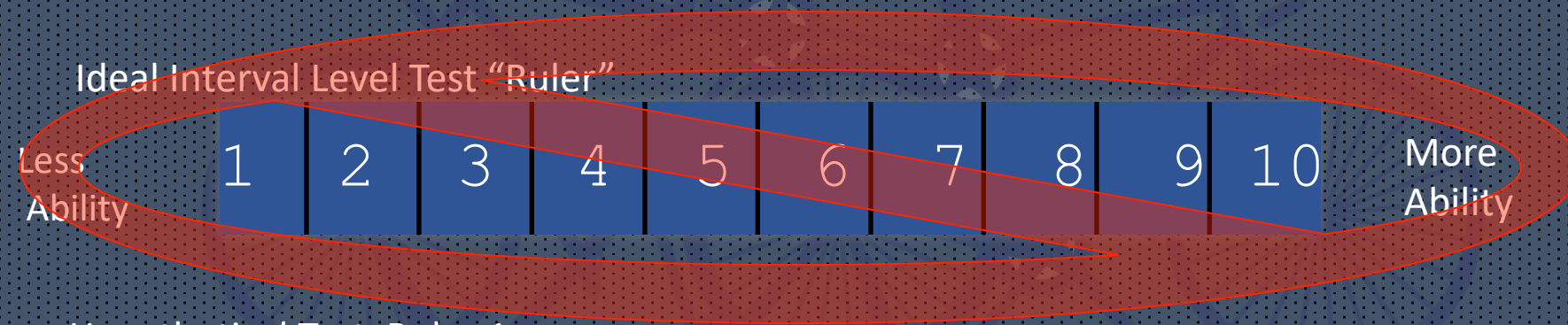
Hypothetical Test Ruler 1



Hypothetical Test Ruler 2



PROBLEM: There is NO external norm to validate the measurement instrument



Hypothetical Test Ruler 1

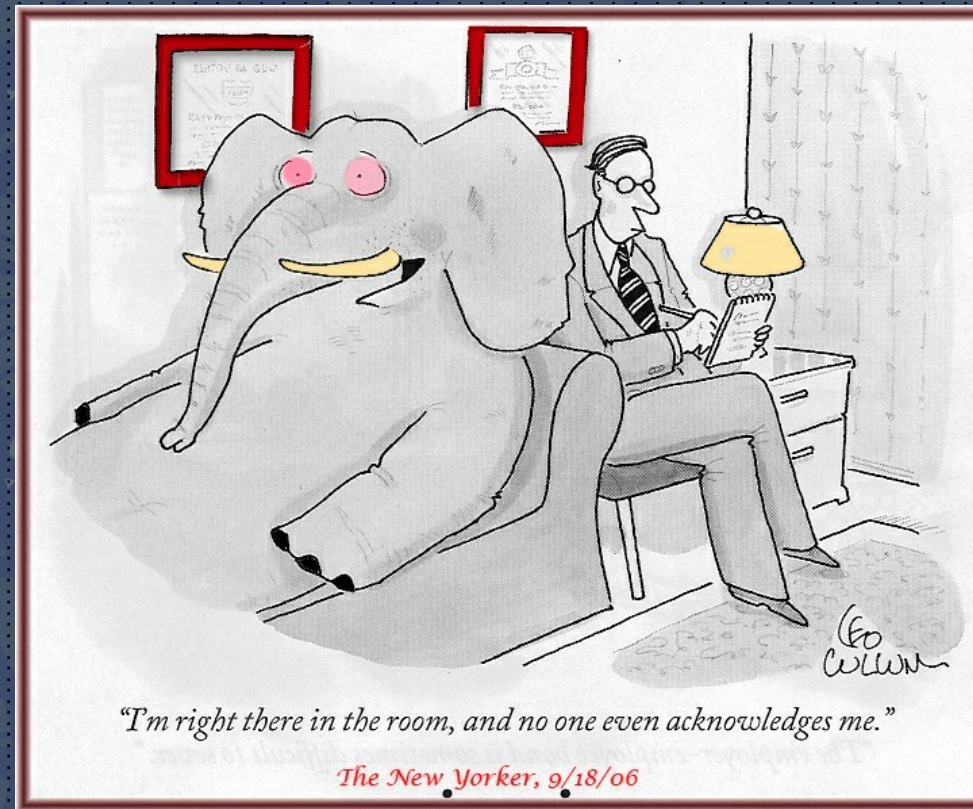


Hypothetical Test Ruler 2



Most Educational Tests

- Are NOT interval (though everyone pretends they are)
- Are probably more ordinal than anything else



Constructing Measure Recap

- What is your context?
- What is your purpose?
 - Criterion- or Norm- Referenced
- Has it been defined unidimensionally?
- What have you done to account for local independence?

With LCTLs, what do you want to measure?

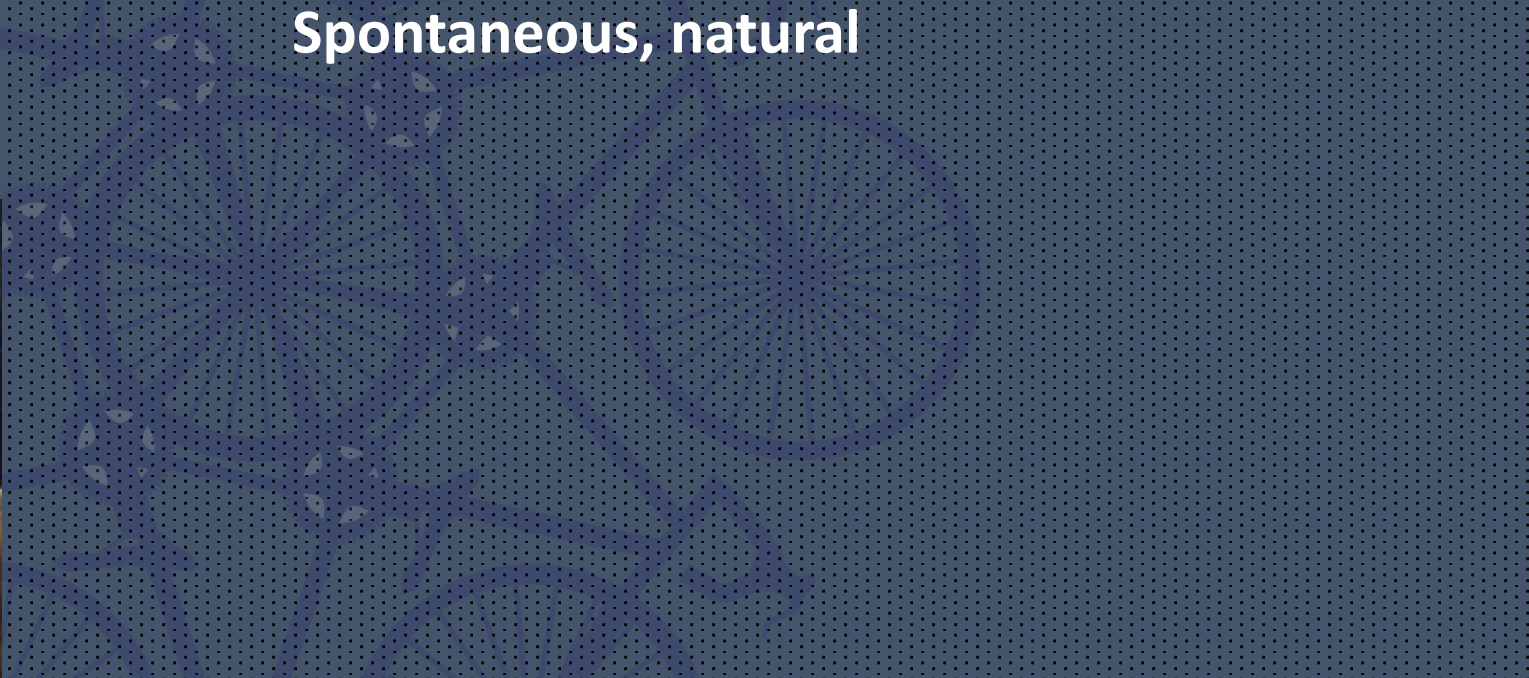
- Knowledge
- Culture
- Literature
- Film
- Vocabulary
- Grammar
- Achievement
- Performance
- Proficiency

Performance Vs. Proficiency

Rehearsed, Practiced



Spontaneous, natural



LCTL Construct Map Activity

- What is proficiency?
 - Functional language ability
 - The ability to USE language for real world purposes
 - Not what a learner knows about language
 - What users of a language can and cannot do

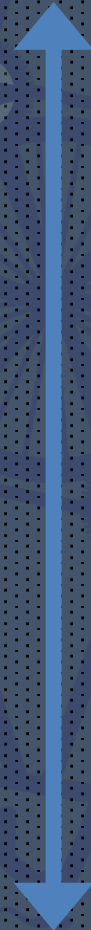
LCTL Construct Map activity

People

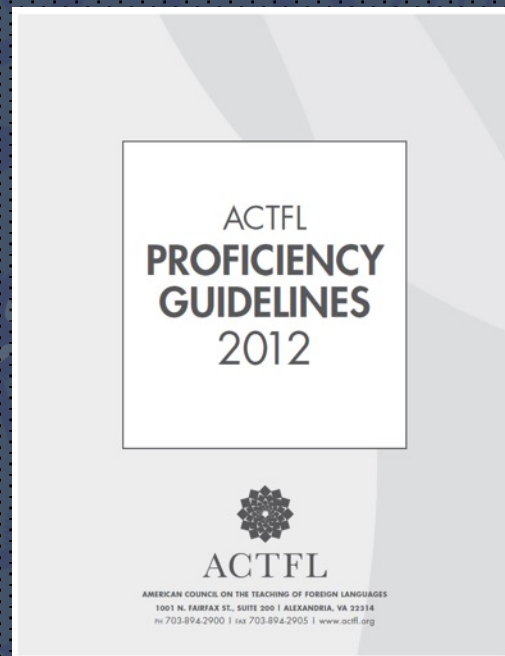
- Think of people you know with three different language proficiency levels in your LCTL.
- Describe how they are different than each other.

Items/Tasks

- Think of tasks in your LCTL with three different language difficulty levels in your LCTL.
- Describe how they are different than each other.



ACTFL Proficiency Guidelines 2012



www.actflproficiencyguidelines2012.org

Schedule/Attend an ACTFL Workshop

- <https://www.actfl.org/assessment-professional-development/professional-development-workshops/proficiency-assessment-workshops>

The screenshot shows the ACTFL (American Council on the Teaching of Foreign Languages) website. The top navigation bar includes links for STORE, COMMUNITY, SIGN UP, and LOG IN. Below this is a search bar. The main navigation bar highlights 'ASSESSMENT & PD' among other categories like ABOUT ACTFL, CONVENTION & EXPO, MEMBERSHIP, PUBLICATIONS, ADVOCACY, and NEWS. The left sidebar lists various resources under 'ASSESSMENT & PD', including 'Workshops' and 'Proficiency & Assessment Workshops'. The main content area is titled 'PROFICIENCY & ASSESSMENT WORKSHOPS' and contains an introductory paragraph, a list of bullet points describing the benefits and topics of the workshops, and a call-to-action button labeled 'ARRANGE A WORKSHOP'. At the bottom, there is a contact information line for further questions.

ACTFL AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES

STORE COMMUNITY SIGN UP LOG IN

Search...

ABOUT ACTFL CONVENTION & EXPO MEMBERSHIP **ASSESSMENT & PD** PUBLICATIONS ADVOCACY NEWS

ASSESSMENT & PD

- ▶ ACTFL Center for Assessment, Research & Development
-
- Quality Assurance
-
- Language Testing International
-
- Research Priorities
-
- Internship & Scholars Program
-
- ▶ Tester & Rater Certifications
-
- Virtual Learning Modules
-
- ▼ **Workshops**
- ▼ **Proficiency & Assessment Workshops**
- ▶ ACTFL Proficiency & Performance Institute
- Curriculum, Instruction, and Classroom Assessment
-
- ▶ Career Resources

PROFICIENCY & ASSESSMENT WORKSHOPS

There are many choices institutions face when choosing the best professional development for their educators. With all the options available, why should you choose an ACTFL workshop?

- ACTFL Workshops are customizable for any type of educational institution or organization
- Our workshop facilitators come from a wide range of linguistic, professional, and educational backgrounds, making it possible to find the appropriate expert to work with your faculty
 - In addition to experts with K-12, higher education, immersion, and applied linguistics backgrounds, we are able to accommodate groups of LCTL (Less Commonly Taught Language) teachers, native and indigenous language communities, and international schools
- Our [workshop topics](#) range from 1-day familiarizations on our assessments (such as OPI and AAPPL) to multi-day workshops on assessing proficiency in written or oral modes of communication
 - All workshops address how ACTFL guidelines and assessment concepts can be applied in a daily classroom setting

Know which workshop you'd like to host?

ARRANGE A WORKSHOP

Have further questions or don't see the topic you need listed? Please contact workshops@actfl.org for more information.