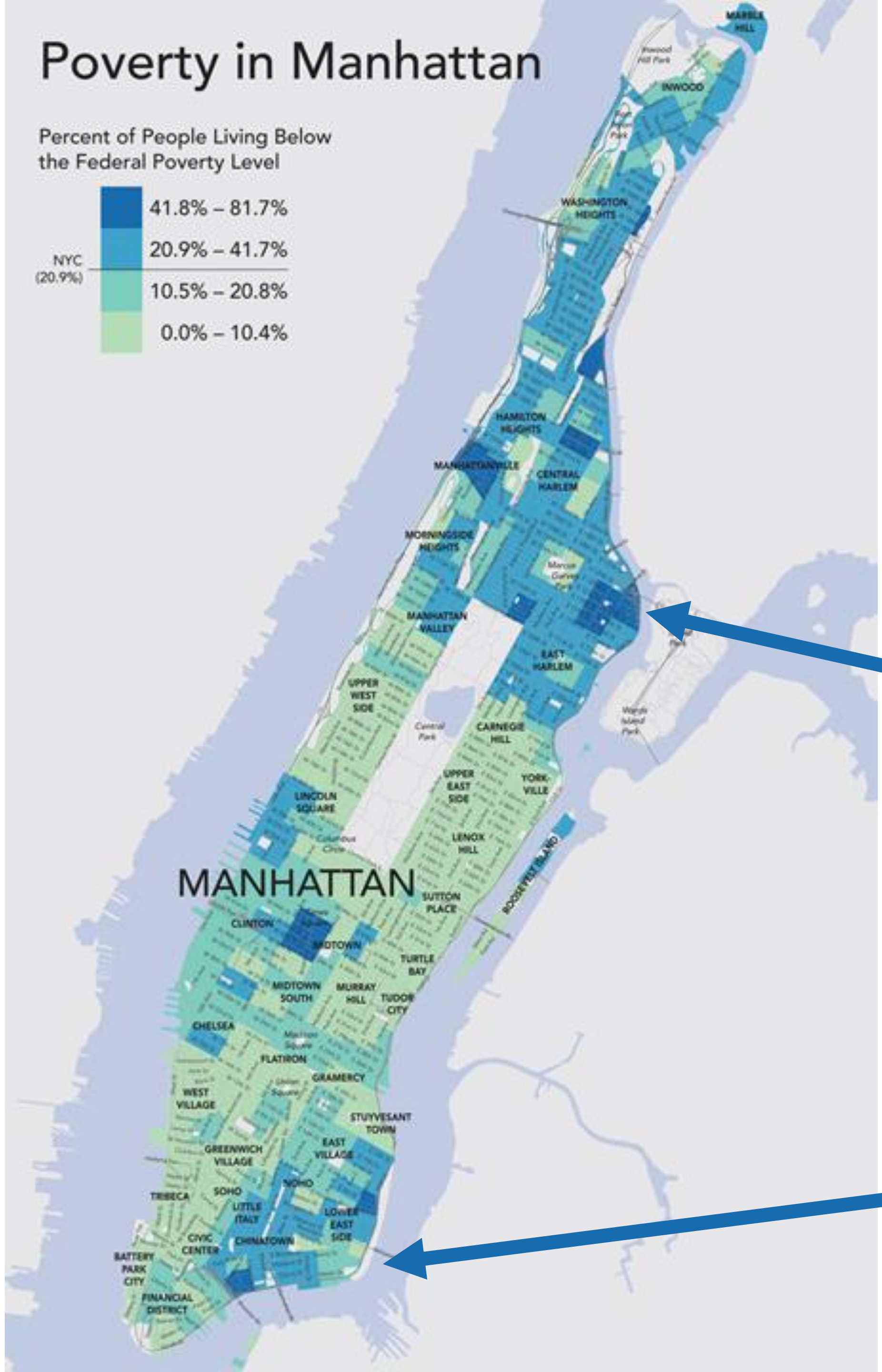
MEASURMENT

Theory



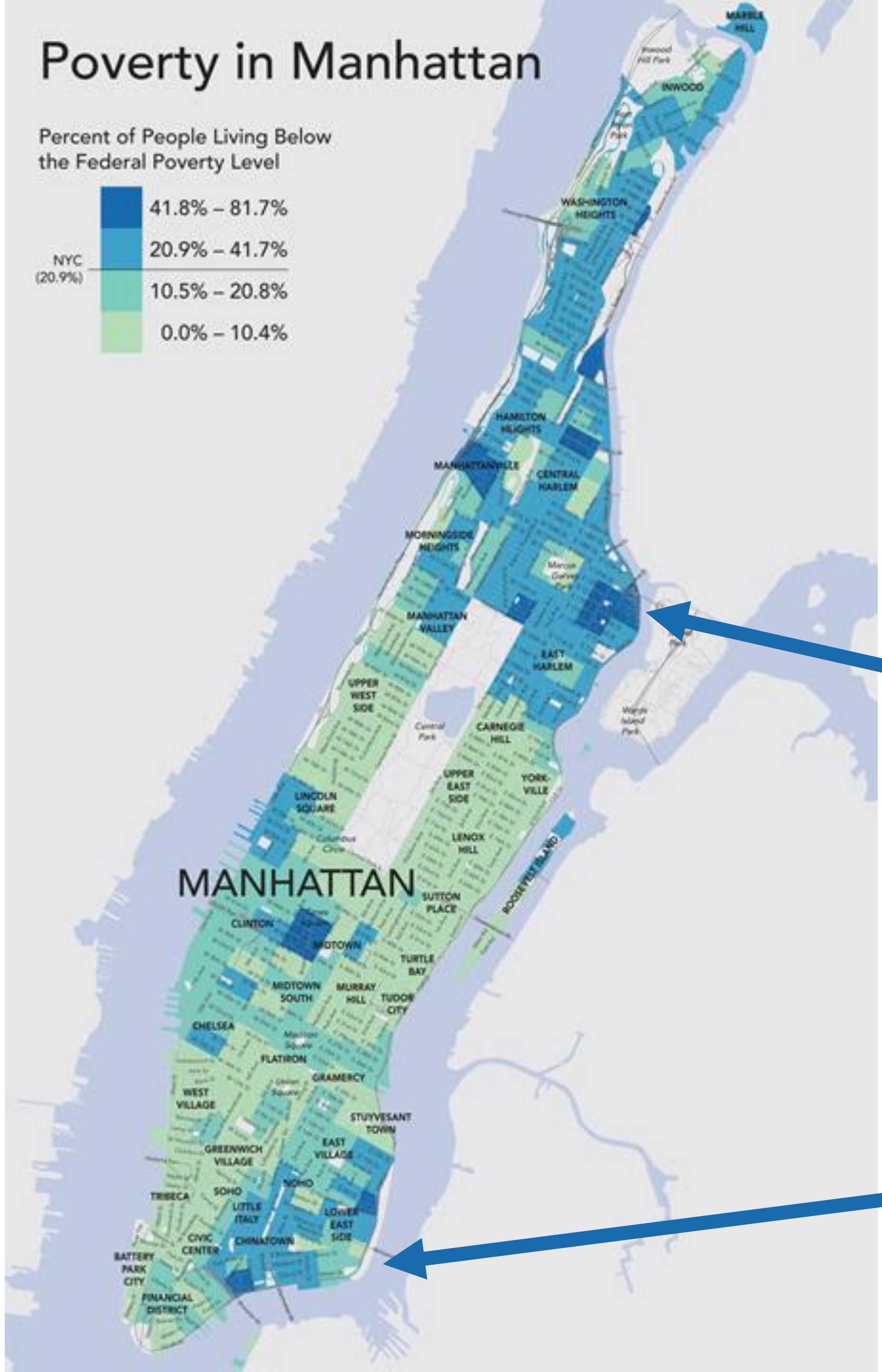
WHAT DOES THE POVERTY RATE MEASURE?

HARLEM

(40% – 80% poverty)

CHINATOWN

(40% – 80% poverty)



WHAT DOES THE POVERTY RATE TELL US ABOUT THE COMMUNITY? WHAT ARE WE MEASURING?

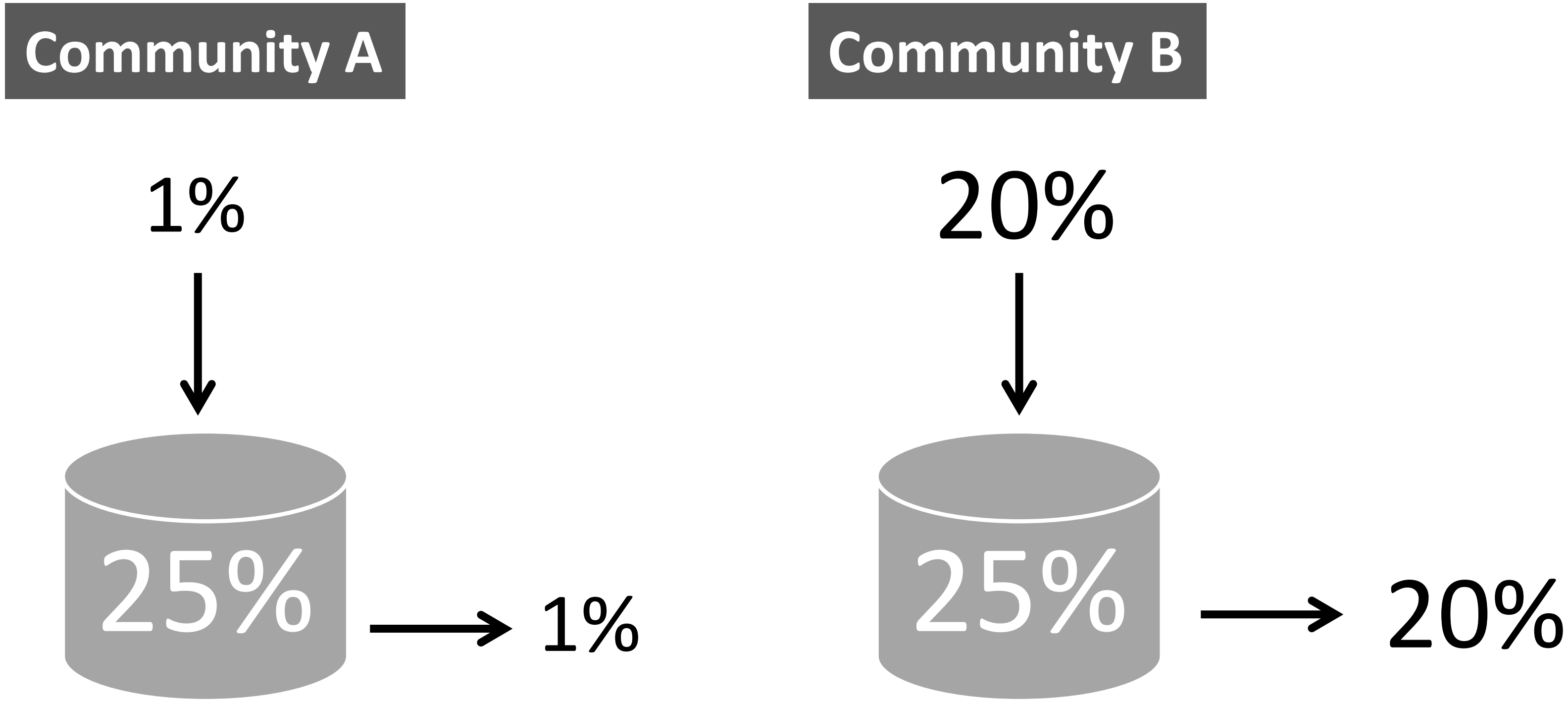
HARLEM

(40% – 80% poverty)

CHINATOWN

(40% – 80% poverty)

The Poverty Rate Hides a Lot:



Community A

1%

25%

1%

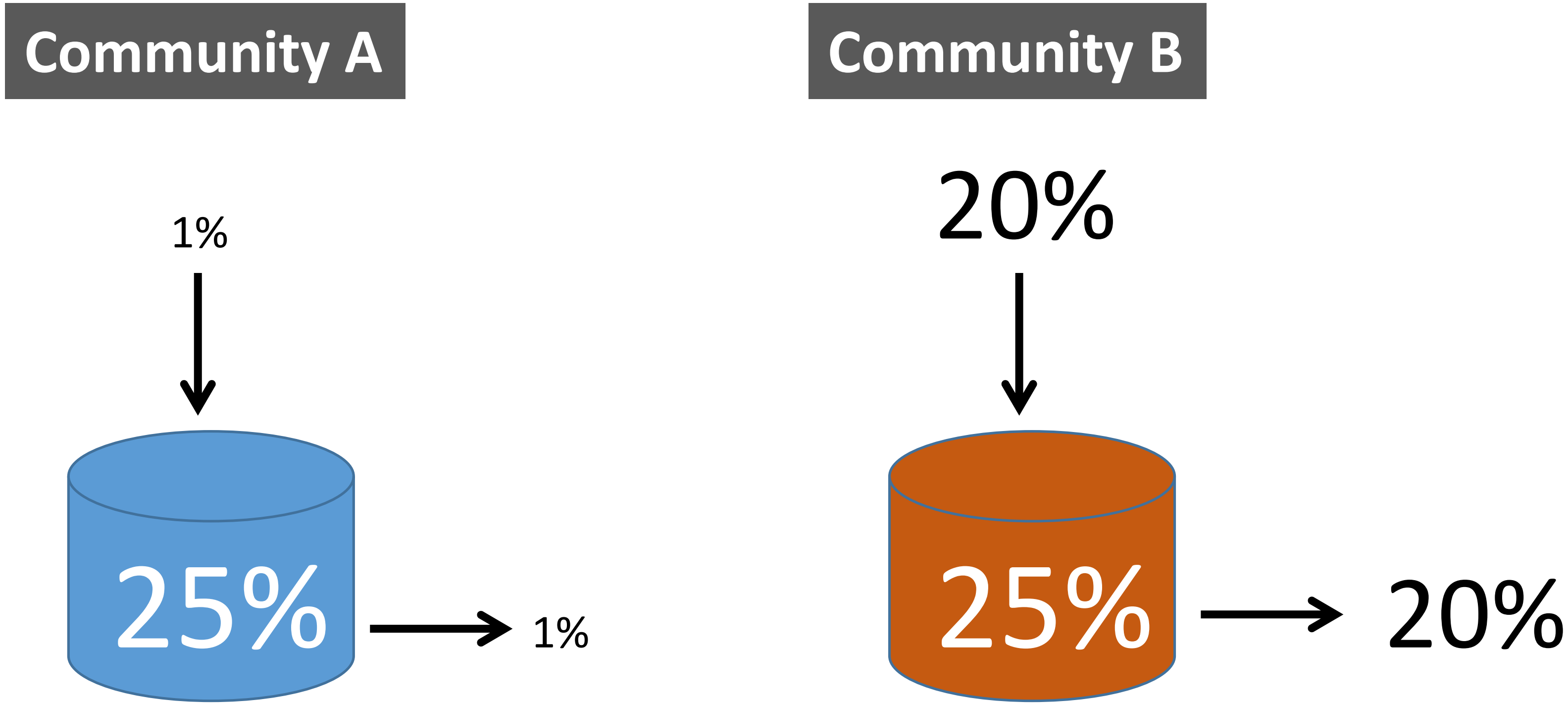
Community B

20%

25%

20%

The Poverty Rate Hides a Lot:



Community A

1%

25%

1%

Community B

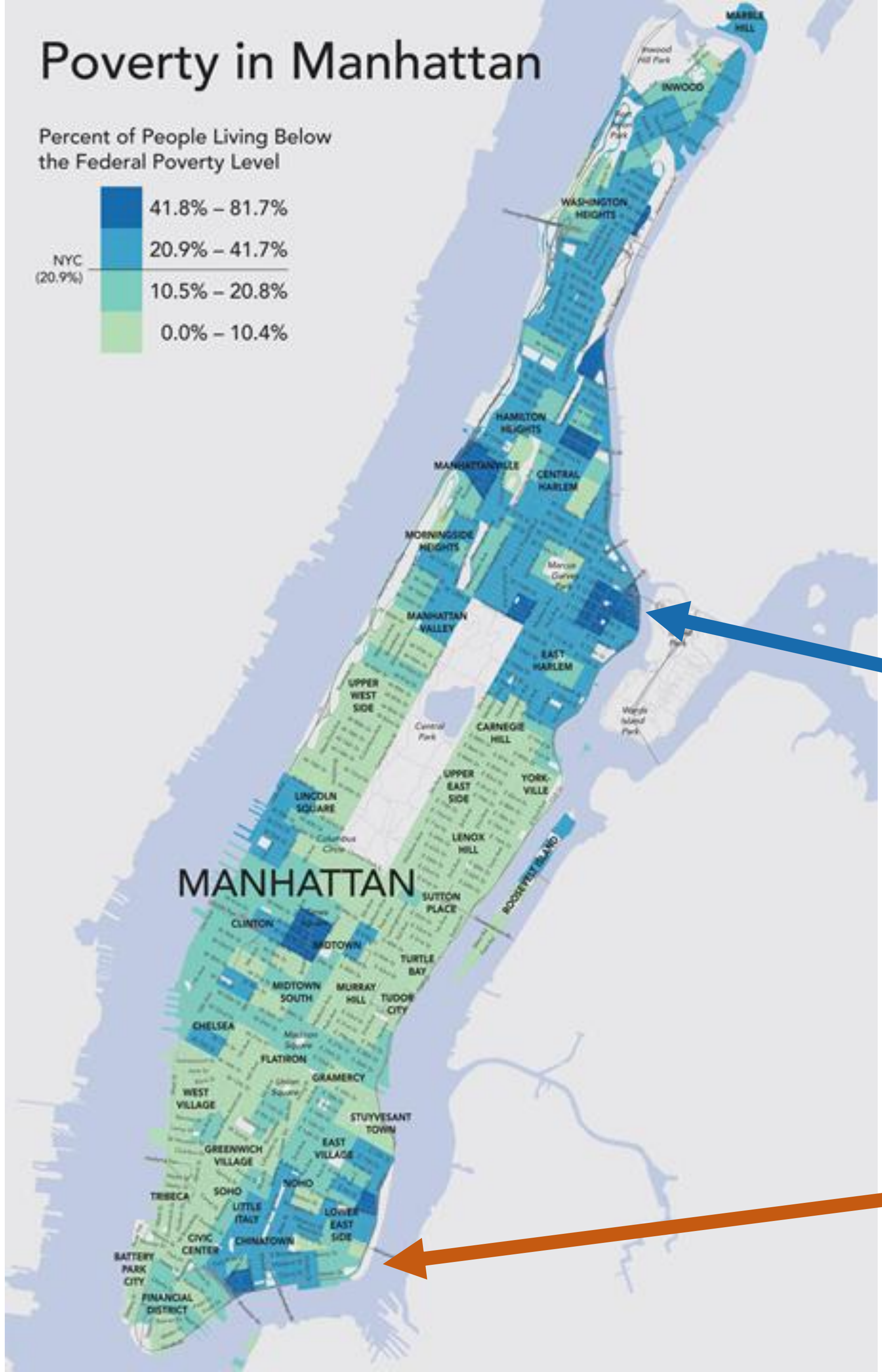
20%

25%

20%

24% Chronic Poverty

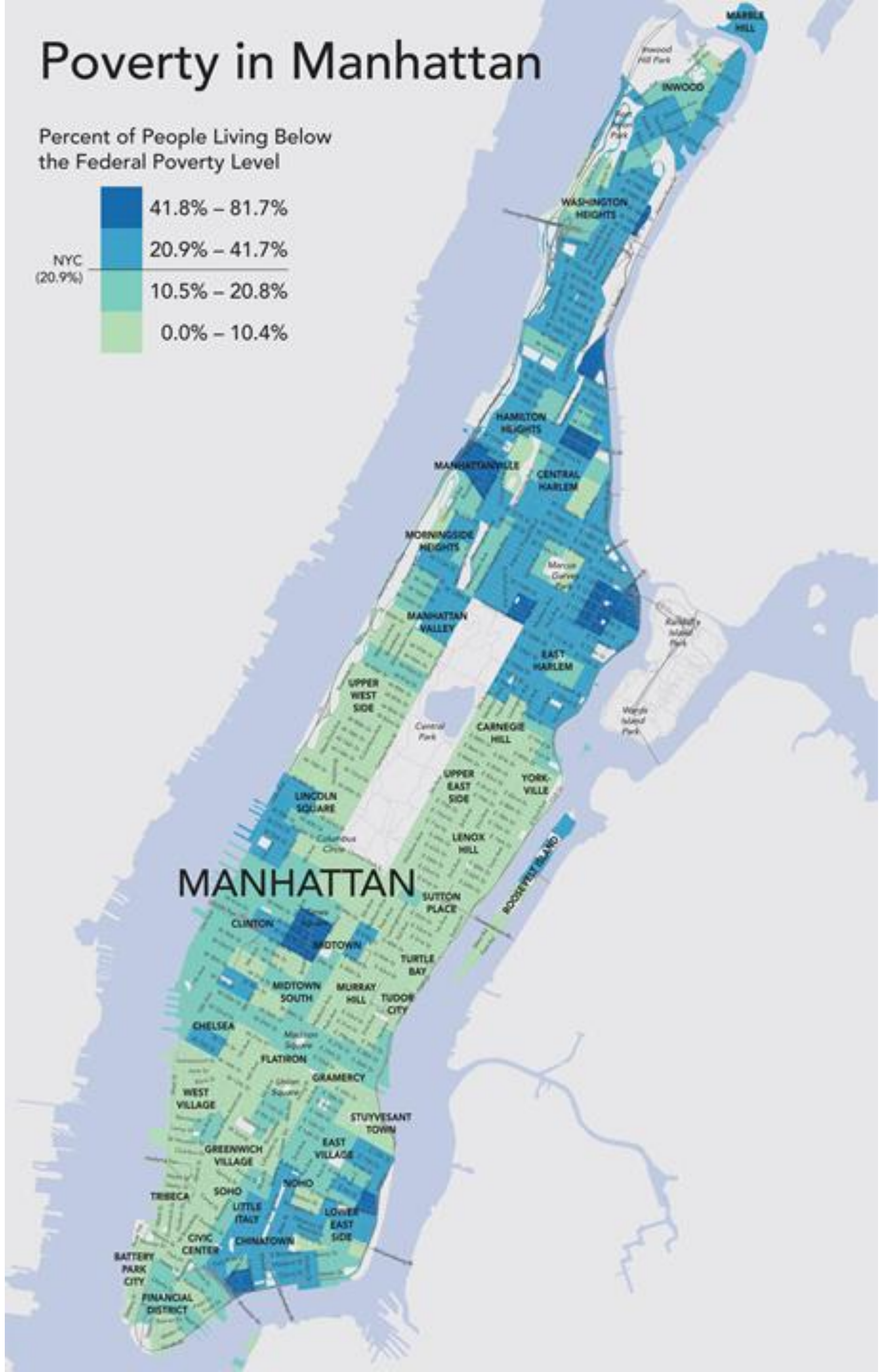
5% Chronic Poverty



NEIGHBORHOODS WITH SIMILAR RATES CAN BE VERY DIFFERENT

Mostly people born in the US. High rates of inter-generational poverty.

Many new immigrants that have few financial assets but strong social capital. Children have high mobility.



WHAT DOES THE POVERTY RATE ACTUALLY MEASURE?

- Lack of money?
- Lack of character, morals?
- Lack of economic opportunity?
- Limited access to healthcare?
- Lack of education?
- Lack of mobility?
- Position in a caste?

- Is a college student
  on a fixed budget poor?

IF YOU COULD DEFINE A FEDERAL MEASURE OF POVERTY, HOW WOULD YOU CREATE A MEANINGFUL SCORE?

WHAT TYPE OF POVERTY WOULD YOUR INDEX MEASURE?

IF YOU COULD DEFINE A NEW FEDERAL MEASURE OF POVERTY, HOW WOULD YOU IMPROVE IT?

Dimensions:

1. Financial capital
2. Human capital
3. Social capital
4. Physical health
5. Public goods (if you have good parks, free libraries, public art do you need money?)

Theory of

MEASUREMENT

TYPES OF MEASURES:

Direct Measures: # of Windshields Installed by a Factory Worker

Markers/Predictors: Direct measures that serve as proxies for correlates that are harder to measure

Latent Constructs: Intelligence (IQ test), Depression (Survey), Health (Survey)

VALID AND RELIABLE INSTRUMENTS:

What do we mean by an "**instrument**"?

What is measurement **validity**?

What is measurement **reliability**?

INSTRUMENTS:

Direct Measures ← instruments are microscopes, spectrometers, and scales

Latent Constructs ← instruments are survey questions, observational protocols for coding data, standardized exams

Instrument

RELIABILITY

VALID AND RELIABLE INSTRUMENTS:

Latent Construct: **Happiness**

Instrument: **Oxford Happiness Questionnaire:**

1. I don’t feel particularly pleased with the way I am. (R) ______
2. I am intensely interested in other people. ______
3. I feel that life is very rewarding. ______
4. I have very warm feelings towards almost everyone. ______
5. I rarely wake up feeling rested. (R) ______

1 = strongly disagree
2 = moderately disagree
3 = slightly disagree
4 = slightly agree
5 = moderately agree
6 = strongly agree

VALID AND RELIABLE INSTRUMENTS:

Latent Construct: **Good Dancer**

Four-Item Survey Instrument:

1. Other people would say I am a good dancer.
2. I am athletic.
3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

VALID AND RELIABLE INSTRUMENTS:

Latent Construct: **Good Dancer**

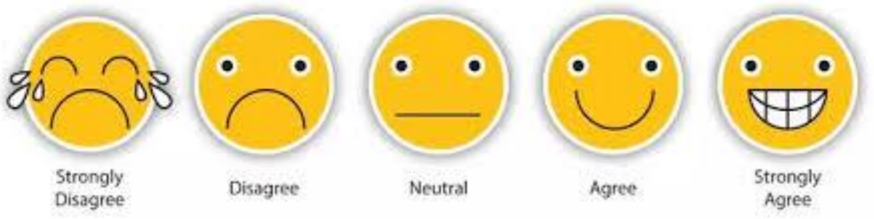
1. Other people would say I am a good dancer [0-4].
2. I am athletic [0-4].
3. I am the first one on the dance floor [0-4].
4. My dance moves have been compared to Drake [0-4].

Good Dancer Measure:
Scale of 0 to 16

Higher is Better

Scale = Instrument

0 1 2 3 4



IS THE SCALE VALID?

Latent Construct: **Good Dancer**

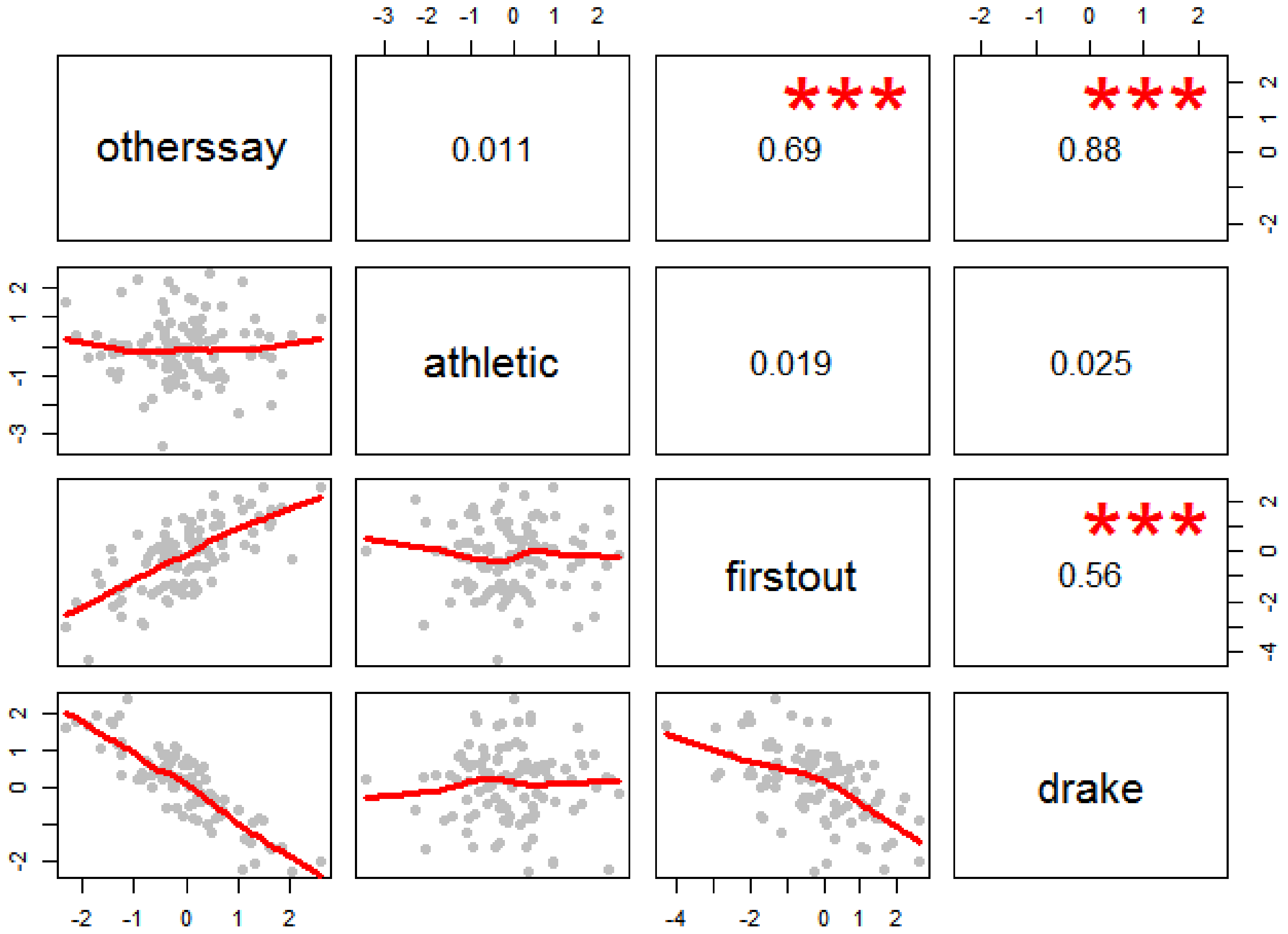
Four-Item Survey Instrument:

1. Other people would say I am a good dancer.
2. I am athletic.
3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

Do we think these items will all measure the same latent construct?

CORRELATION STRUCTURE

Which of these are measuring the same thing?



IS THE SCALE RELIABLE?

Latent Construct: **Good Dancer**

Four-Item Survey Instrument:

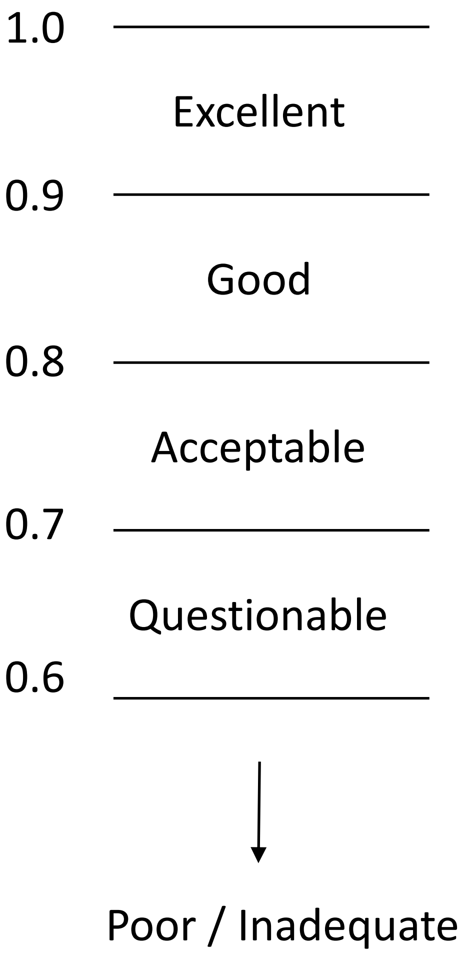
1. Other people would say I am a good dancer.
2. I am athletic.
3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

How consistently do these questions measure the same latent construct?

Cronbach's alpha is a measure of internal consistency, that is, how closely related a set of items are as a group.

It is considered to be a measure of scale reliability.

CRONBACH'S ALPHA SCORE MEASURE OF RELIABILITY [0, 1]



1.0

Excellent

0.9

Good

0.8

Acceptable

0.7

Questionable

0.6

Poor / Inadequate

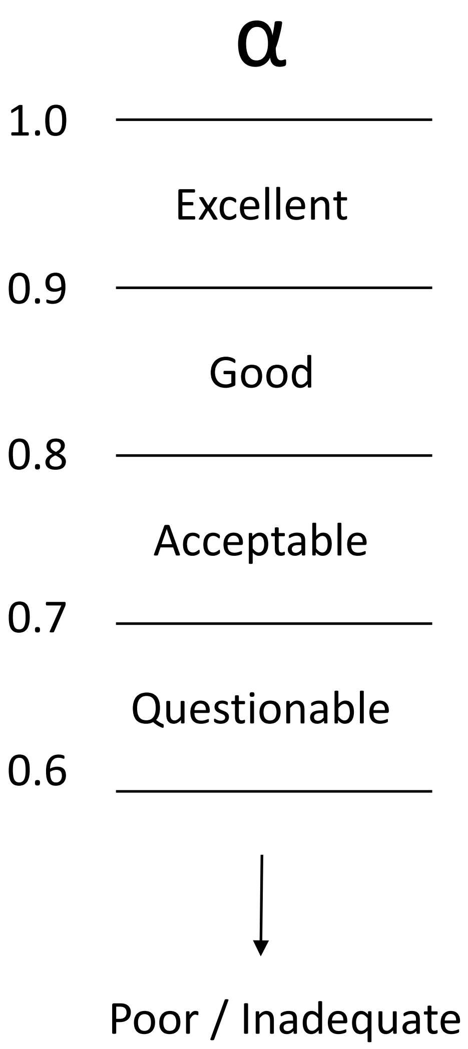
$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

N = number of items
c-bar = average inter-item covariance
v-var = average variance per item

RELIABLE INSTRUMENTS:

1. Other people would say I am a good dancer.
2. I am athletic.
3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

$\alpha = 0.68$



α

1.0

Excellent

0.9

Good

0.8

Acceptable

0.7

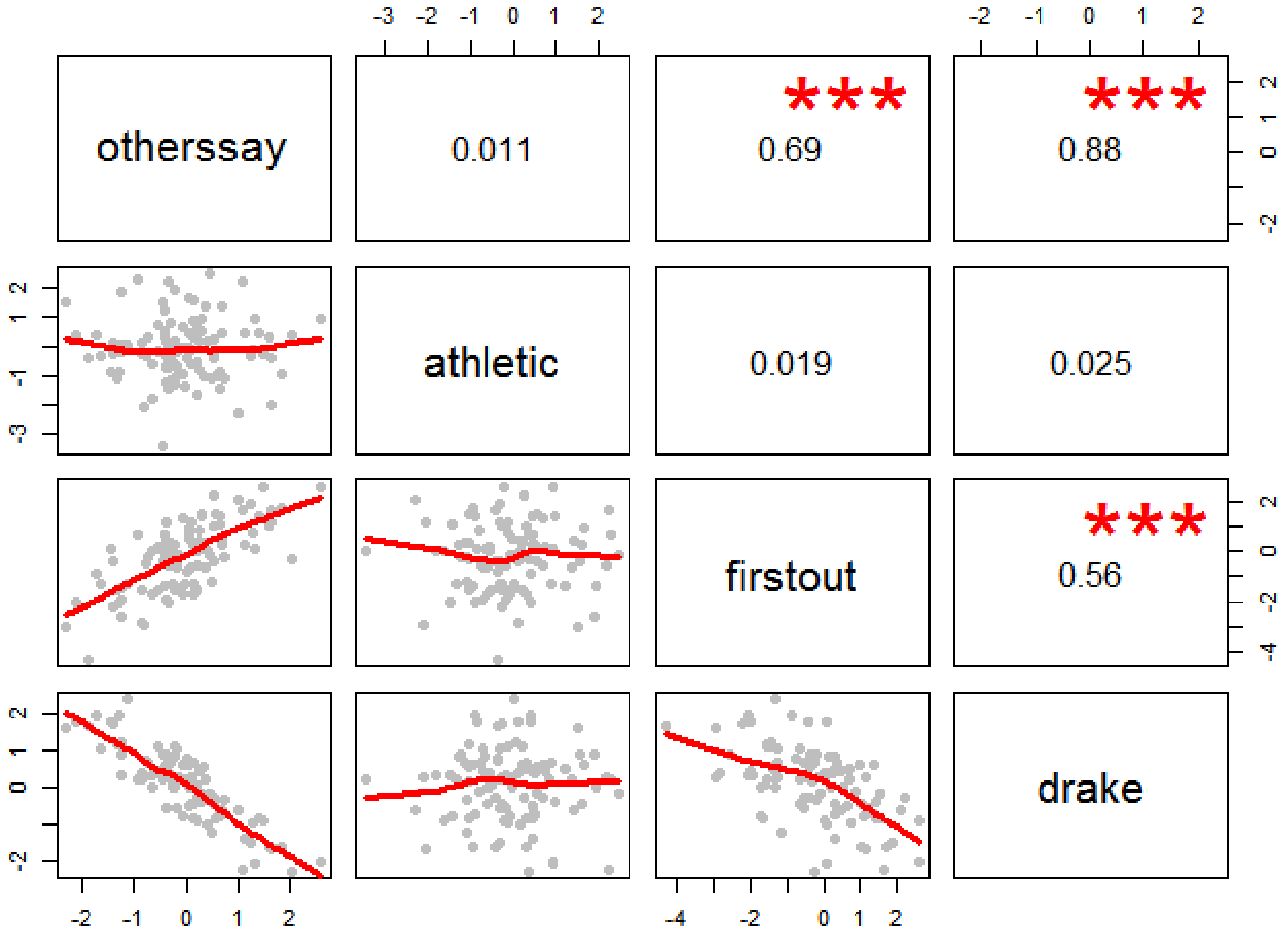
Questionable

0.6

Poor / Inadequate

CORRELATION STRUCTURE

Which of these are measuring the same thing?



RELIABLE INSTRUMENTS:

1. Other people would say I am a good dancer.
2. I am athletic.
3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

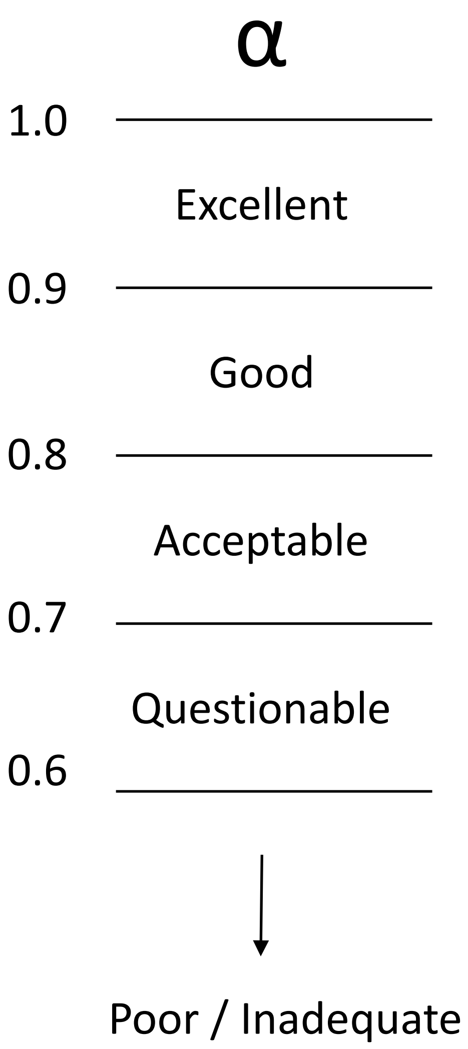
$\alpha = 0.68$

1. Other people would say I am a good dancer.

~~2. I am athletic.~~

3. I am the first one on the dance floor.
4. My dance moves have been compared to Drake.

$\alpha = 0.86$



α

1.0

Excellent

0.9

Good

0.8

Acceptable

0.7

Questionable

0.6

Poor / Inadequate

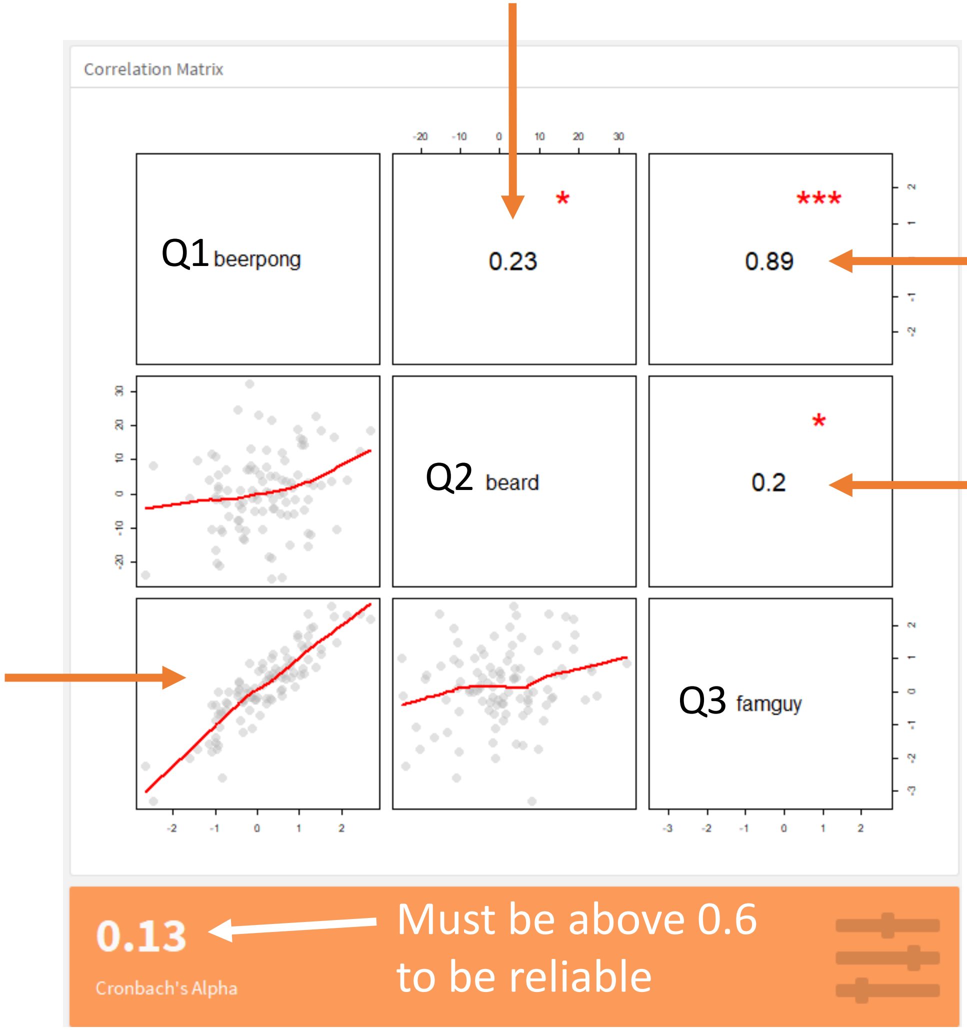
BRO CULTURE

Characterized by:

- Entitlement
- Disregard for others
- Self-destructive behaviors
- Closed-mindedness
- Strong group cohesion

https://www.youtube.com/watch?v=VU3zuK7Zmrk

Correlation between
Q1 and Q2 (1 star=low)


Correlation Matrix

0.13
Cronbach's Alpha

Must be above 0.6
to be reliable

Correlation between
Q1 and Q3 (3 stars=high)

Correlation between
Q2 and Q3 (1 star=low)

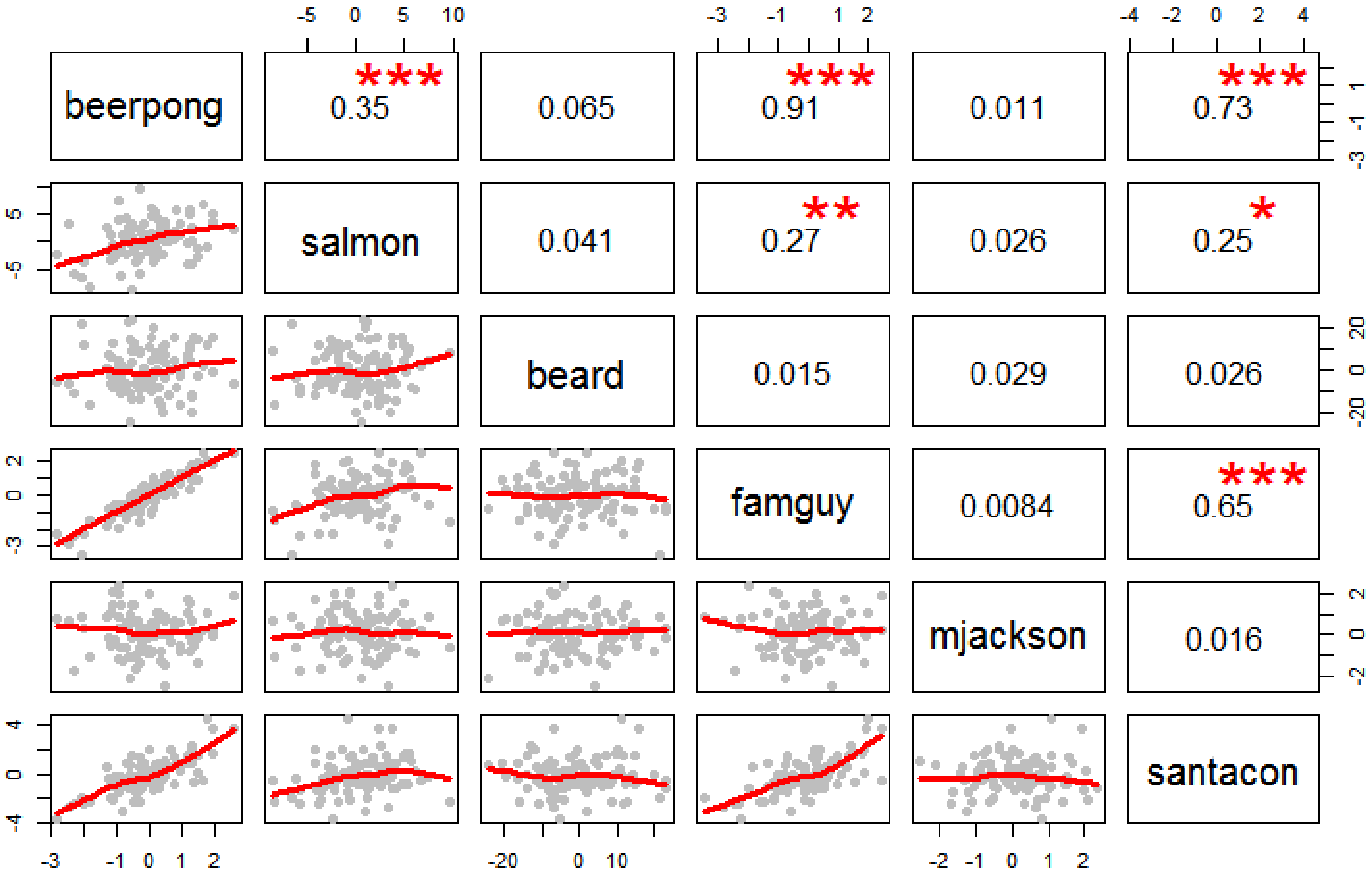
People that like beer pong
also like *The Family Guy*
(responses are
highly-correlated)

SURVEY ITEMS TO IDENTIFY BRO CULTURE:

1. Beer pong is a fun game to play at parties.
2. Salmon is a fine color for shorts.
3. Beards are repulsive.
4. Family Guy is a funny television show.
5. Michael Jackson is one of the best musicians ever.
6. Santacon is a great idea for a festival.

Strongly Disagree ← 1 2 3 4 5 → Strongly Agree

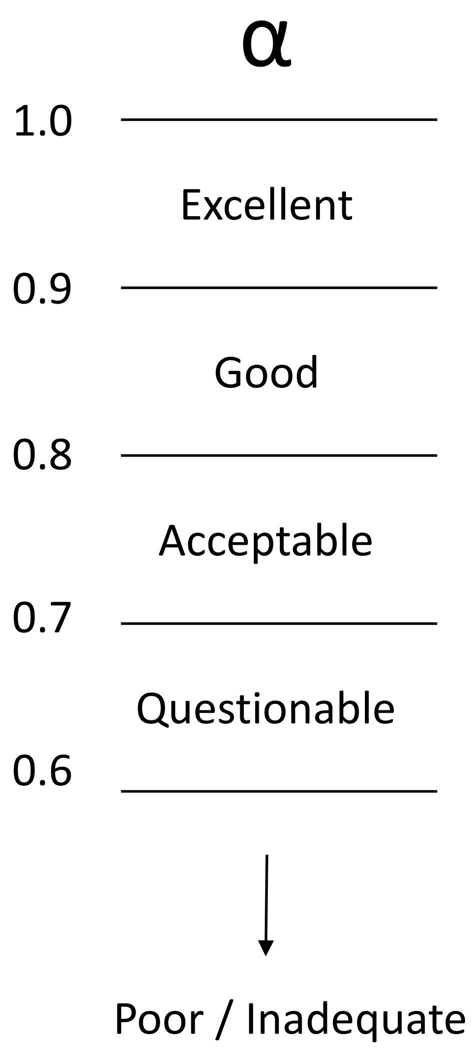
CORRELATION STRUCTURE OF SURVEY ITEMS



CRONBACH'S ALPHA SCORE:

1. Beer pong is a fun game to play at parties.
2. Salmon is a fine color for shorts.
3. Beards are repulsive.
4. Family Guy is a funny television show.
5. Michael Jackson is one of the best musicians ever.
6. Santacon is a great idea for a festival.

$$\alpha = 0.16$$



α

1.0

Excellent

0.9

Good

0.8

Acceptable

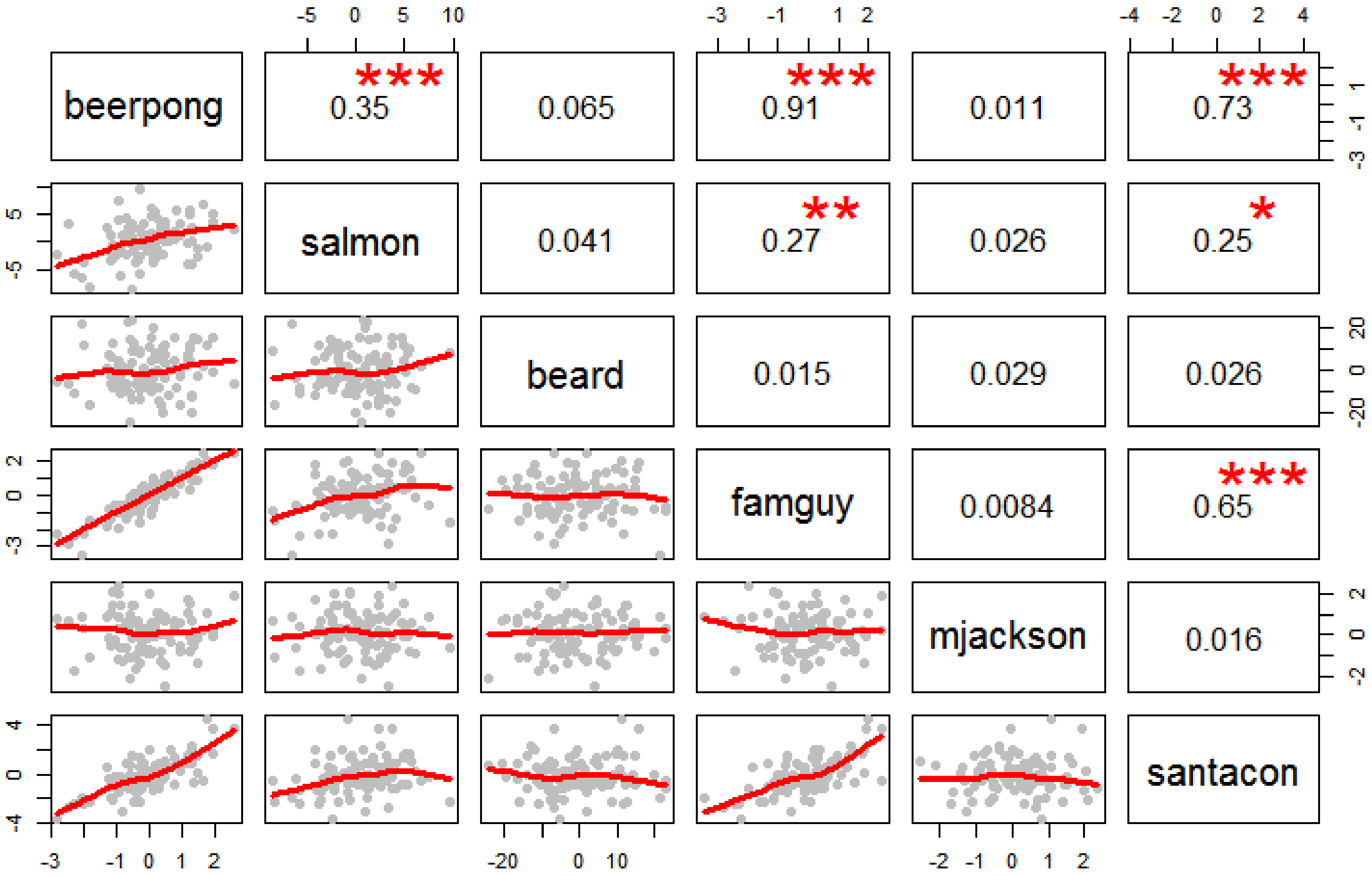
0.7

Questionable

0.6

Poor / Inadequate

CORRELATION STRUCTURE OF SURVEY ITEMS

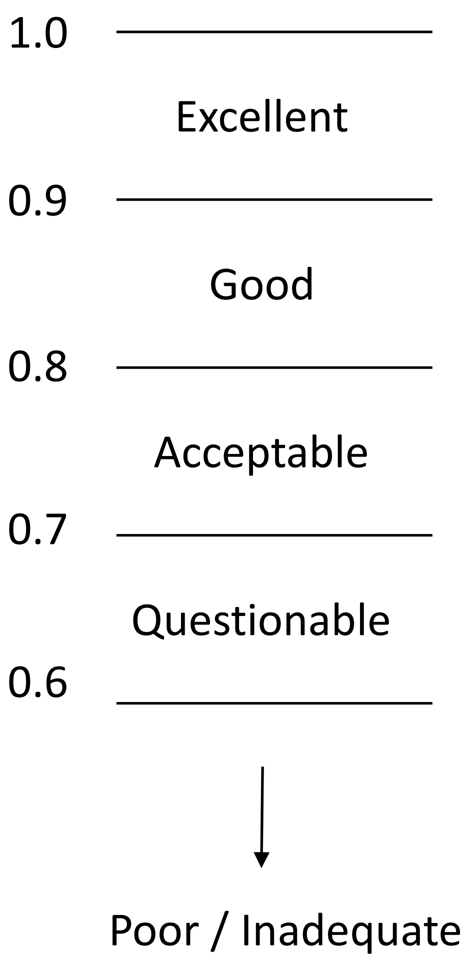


CRONBACH'S ALPHA SCORE:

1. Beer pong is a fun game to play at parties.
2. Salmon is a fine color for shorts.
3. ~~Beards are repulsive.~~
4. Family Guy is a funny television show.
5. ~~Michael Jackson is one of the best musicians ever.~~
6. Santacon is a great idea for a festival.

$\alpha = 0.61$

CRONBACH’S ALPHA SCORE MEASURE OF RELIABILITY [0, 1]



1.0

Excellent

0.9

Good

0.8

Acceptable

0.7

Questionable

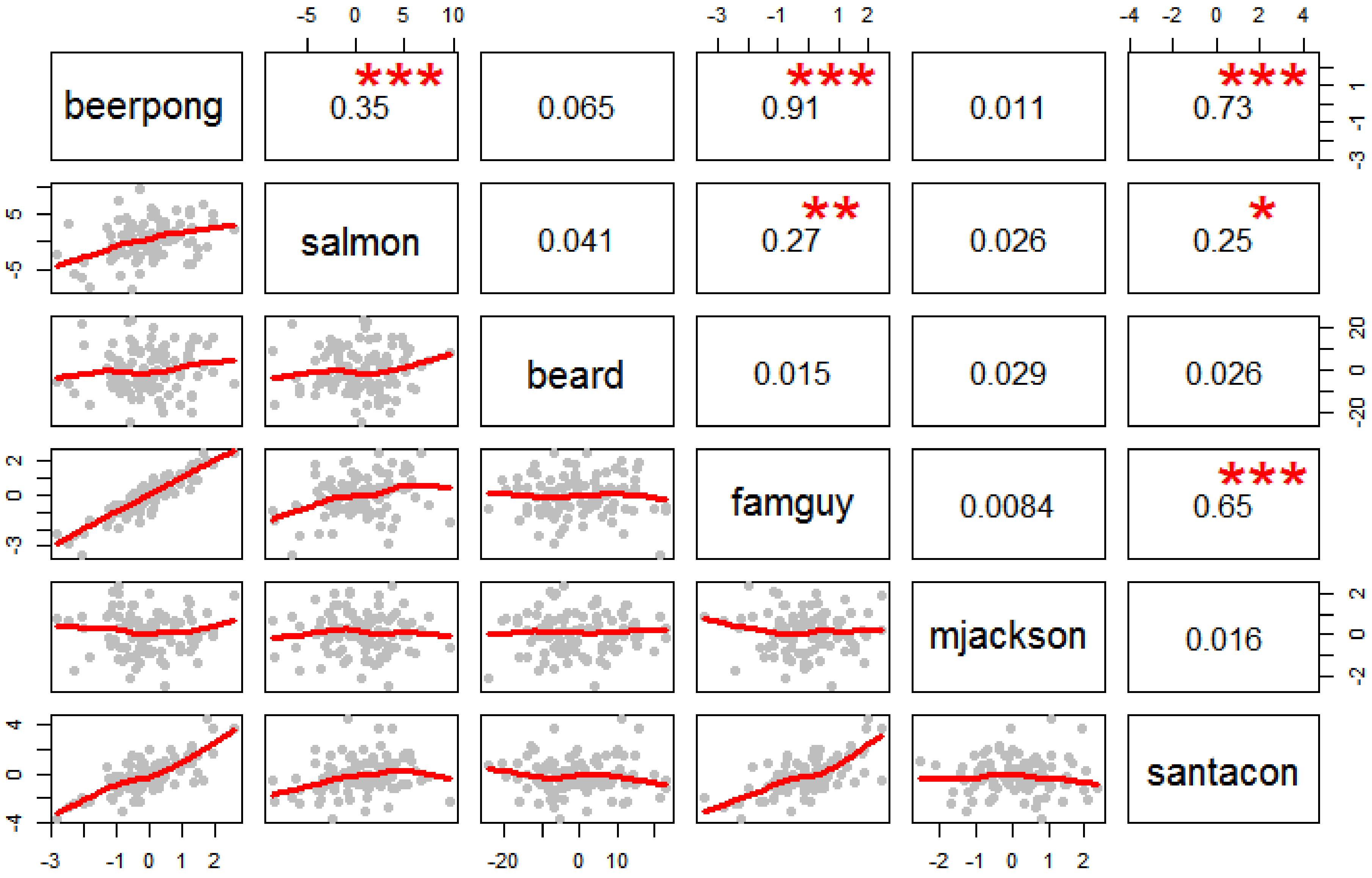
0.6

Poor / Inadequate

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

N = number of items
c-bar = average inter-item covariance
v-var = average variance per item

CORRELATION STRUCTURE OF SURVEY ITEMS



CRONBACH'S ALPHA SCORE:

1. Beer pong is a fun game to play at parties.
2. ~~Salmon is a fine color for shorts.~~
3. ~~Beards are repulsive.~~
4. Family Guy is a funny television show.
5. ~~Michael Jackson is one of the best musicians ever.~~
6. Santacon is a great idea for a festival.

$$\alpha = 0.89$$

EXAMPLES

PROGRESS OUT OF POVERTY INDEX

Grameen Bank: How does the Progress out of Poverty Index (PPI) work?

Unlike other poverty measurement methods, the PPI was designed with the budgets and operations of real organizations in mind; its simplicity means that it requires fewer resources to use.

The PPI is a set of 10 easy-to-answer questions that a household member can answer in 5 to 10 minutes. The questions are simple – “What material is your roof made out of? How many of your children are in school?”

The scored answers provide the likelihood that the survey respondent’s household is living below the national poverty line and other internationally-recognized poverty lines. The PPI is country-specific. There are PPIs for 45 countries

http://www.progressoutofpoverty.org/about-ppi

KRISHNA: STAGES OF PROGRESS

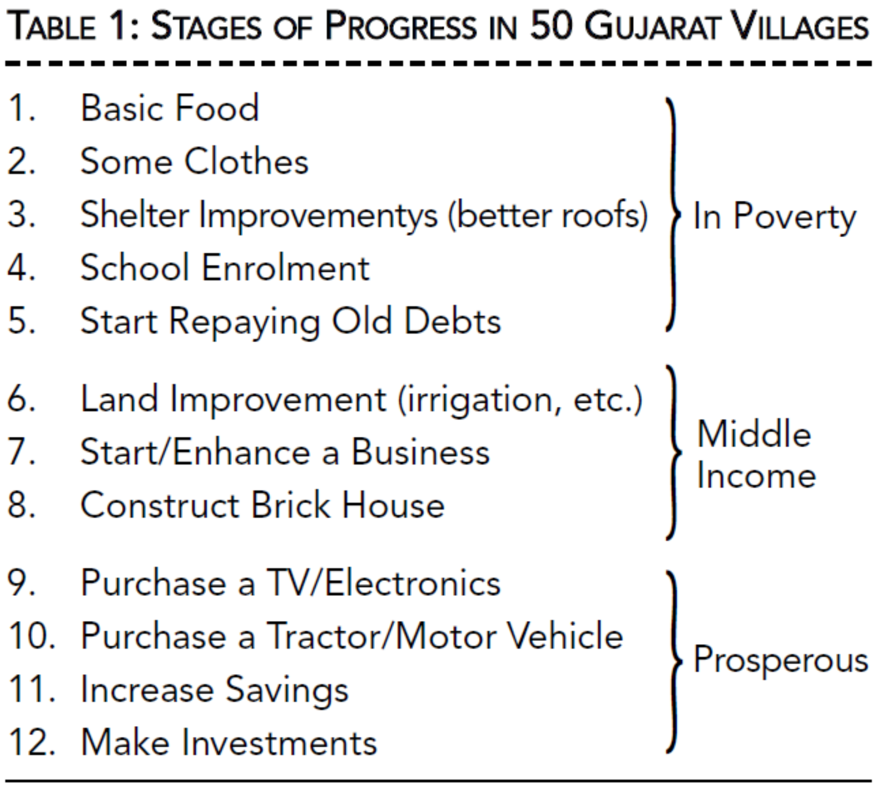
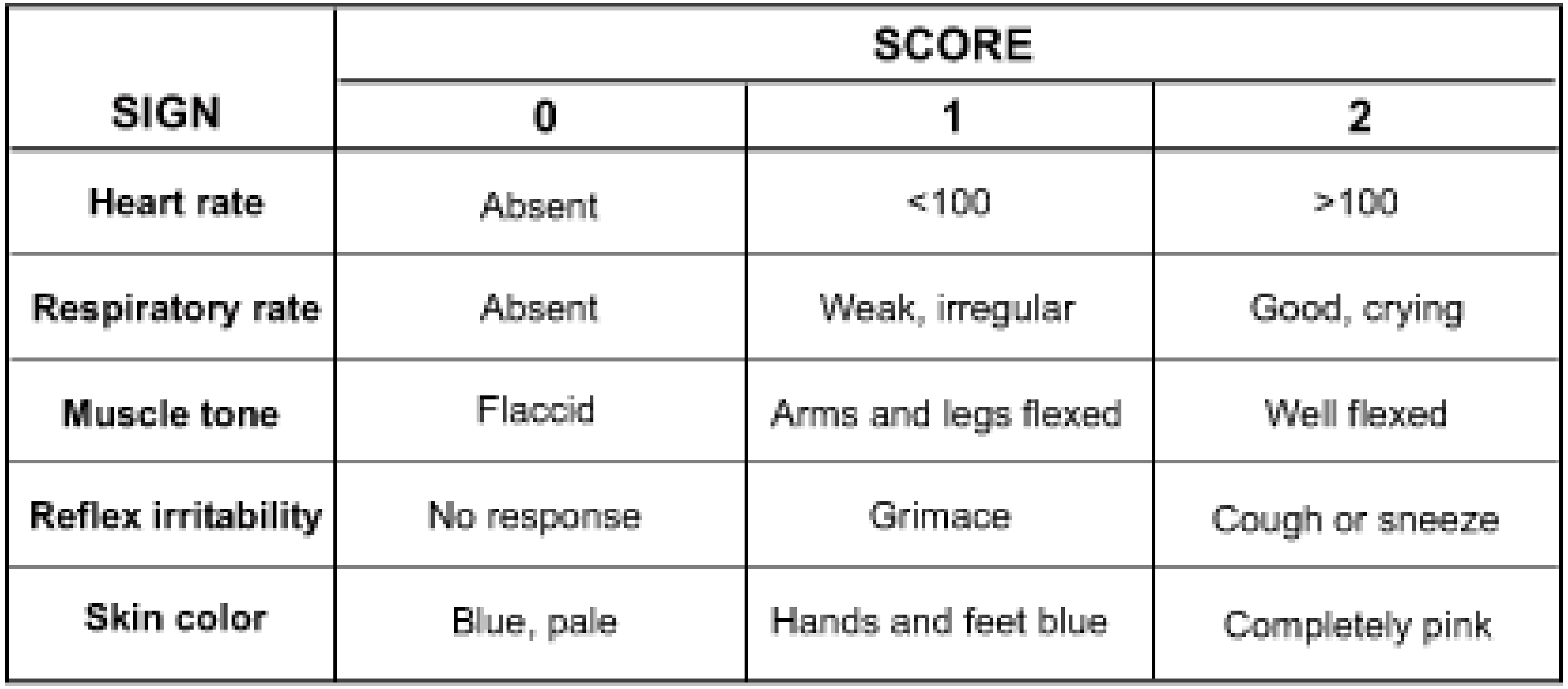


TABLE 1: STAGES OF PROGRESS IN 50 GUJARAT VILLAGES

1.	Basic Food	In Poverty
2.	Some Clothes	
3.	Shelter Improvementys (better roofs)	
4.	School Enrolment	
5.	Start Repaying Old Debts	
6.	Land Improvement (irrigation, etc.)	Middle Income
7.	Start/Enhance a Business	
8.	Construct Brick House	
9.	Purchase a TV/Electronics	Prosperous
10.	Purchase a Tractor/Motor Vehicle	
11.	Increase Savings	
12.	Make Investments	

THE APGAR SCORE



SIGN	SCORE		
	0	1	2
Heart rate	Absent	<100	>100
Respiratory rate	Absent	Weak, irregular	Good, crying
Muscle tone	Flaccid	Arms and legs flexed	Well flexed
Reflex irritability	No response	Grimace	Cough or sneeze
Skin color	Blue, pale	Hands and feet blue	Completely pink

Scale of 0 to 10, measured right after birth, and 5 minutes later.

The politics of

MEASUREMENT

(aside)

MEASUREMENT IS INHERENTLY POLITICAL

Challenges:

1. What if we can’t measure what we care about?
2. Time and money spent on data collection.
3. Perverse Incentives

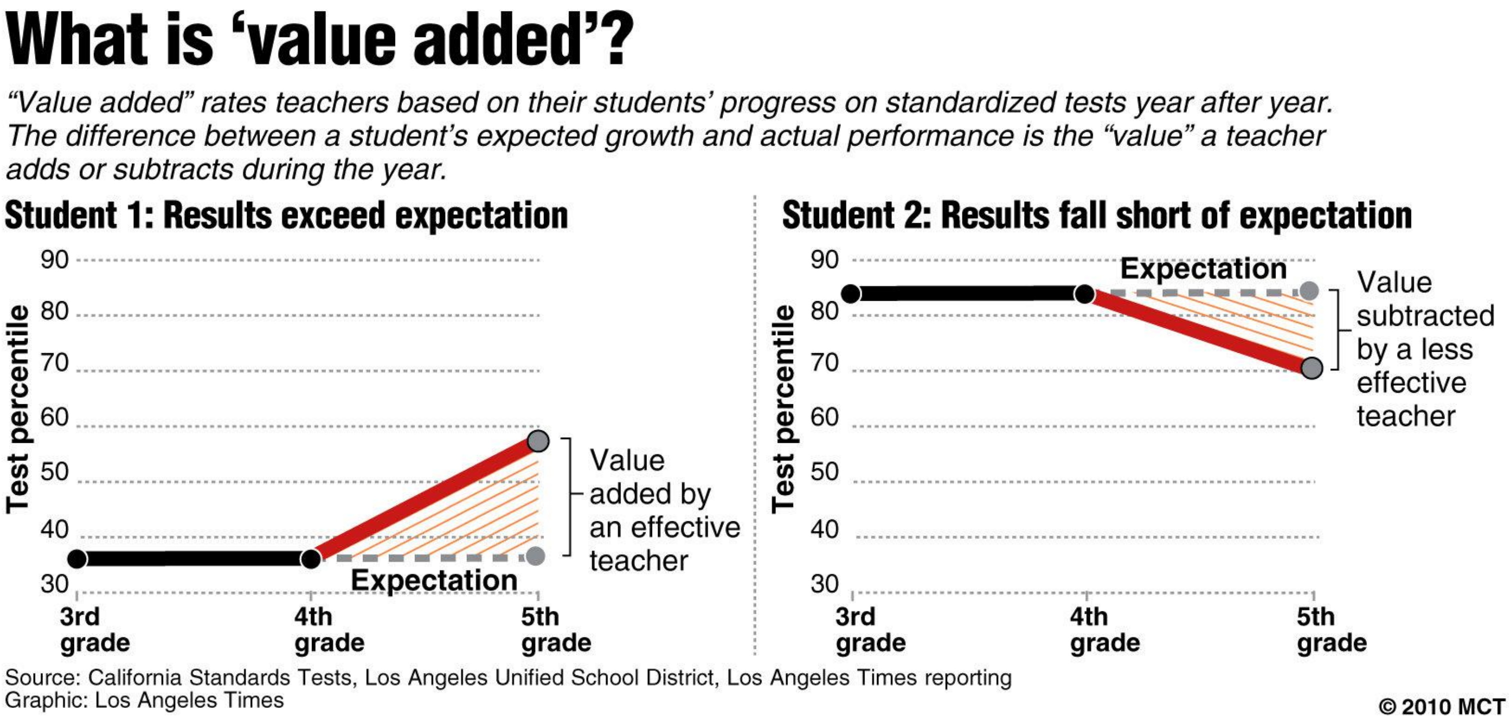
   “Multi-tasking” – what get’s measured gets done, but at a price of other activities.
   Creaming – not providing services to the poorest / hardest cases.

4. Poor Counterfactuals (evaluation capacity)
5. Fatigue from Multiple Stakeholders

HUMANS OF NEW YORK

“I decided to become a teacher because I knew what it was like to grow up poor, and I wanted to help kids in similar circumstances. I didn’t expect it to be easy. But I guess I thought there’d be only one or two kids acting up in class, and everyone else would be paying attention. Instead it’s only one or two kids who actually behave. I’m drained every day. I’ve been teaching for thirteen years. And if it wasn’t for summer break, I’d have quit already.

Forty percent of my job rating is based on standardized testing. It’s the only job I know where your performance is based on how other people behave. I can’t control what’s going on outside my classroom. I can’t control if my kids are from abusive households, or don’t eat breakfast, or can’t get to school on time. But those things affect my rating when they show up in test scores. I need to find a new career where my performance is based on me.”



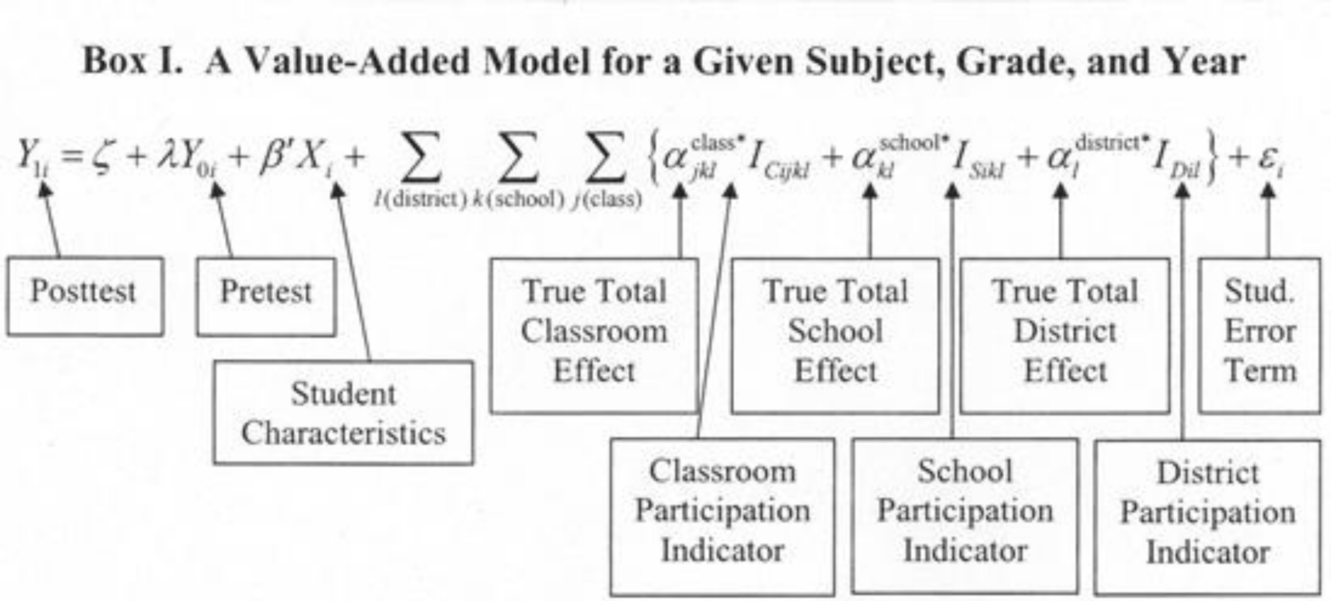
What is 'value added'?

"Value added" rates teachers based on their students' progress on standardized tests year after year. The difference between a student's expected growth and actual performance is the "value" a teacher adds or subtracts during the year.

Source: California Standards Tests, Los Angeles Unified School District, Los Angeles Times reporting
Graphic: Los Angeles Times

© 2010 MCT

See *Weapons of Math Destruction* for examples of how metrics have been weaponized in public agencies

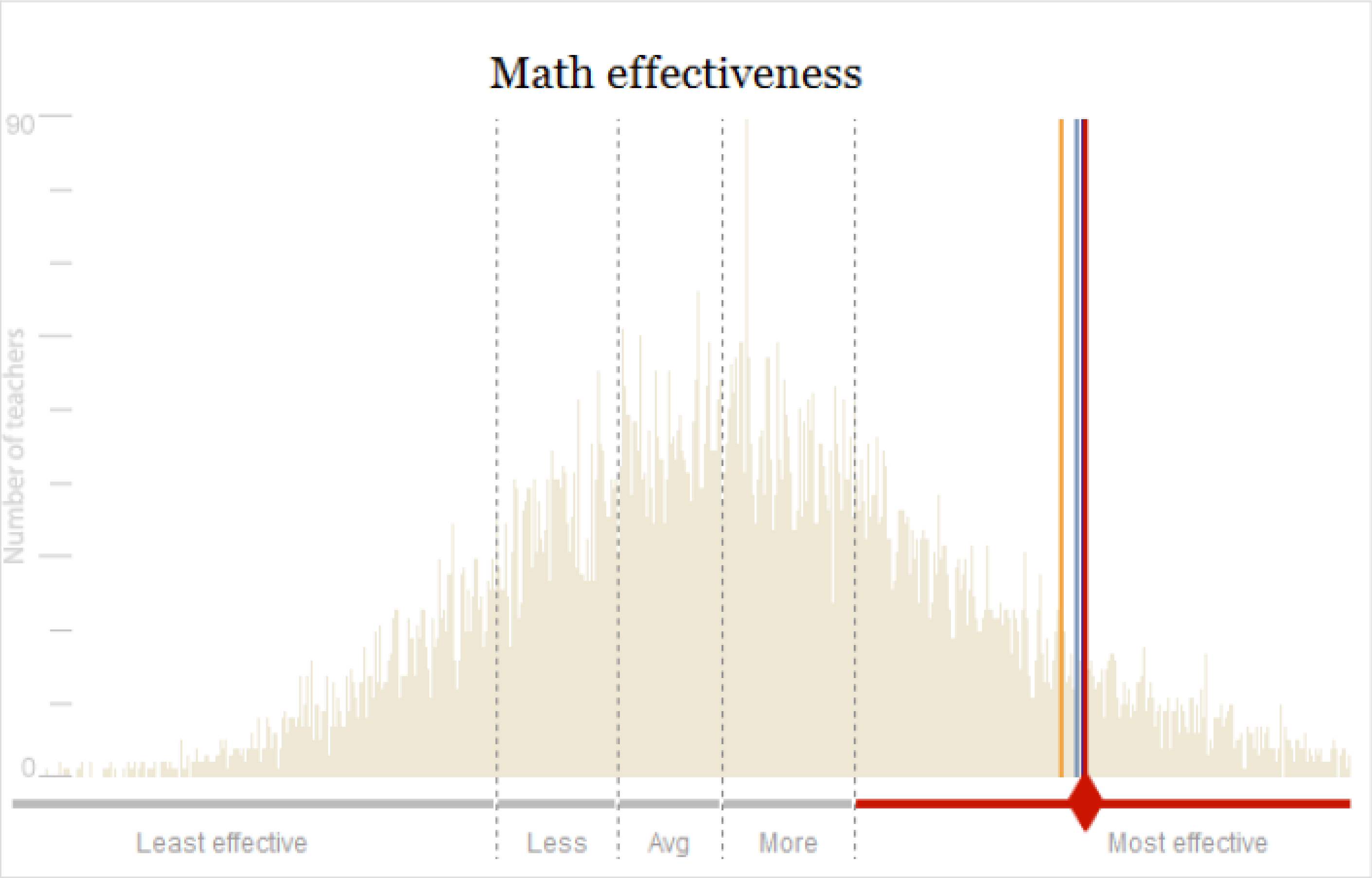


Box I. A Value-Added Model for a Given Subject, Grade, and Year

$$Y_{1i} = \zeta + \lambda Y_{0i} + \beta' X_i + \sum_{l(\text{district})} \sum_{k(\text{school})} \sum_{j(\text{class})} \left\{ \alpha_{jkl}^{\text{class*}} I_{Cijkl} + \alpha_{kl}^{\text{school*}} I_{Sikl} + \alpha_{l}^{\text{district*}} I_{Dil} \right\} + \varepsilon_i$$

Posttest
Pretest
Student Characteristics
True Total Classroom Effect
Classroom Participation Indicator
True Total School Effect
School Participation Indicator
True Total District Effect
District Participation Indicator
Stud. Error Term

COMPARING TEACHER "EFFECTS"



PERFORMANCE OR PUNISHMENT?

The focusing effects of outcome benchmarks, the pressures of competition, the prospects of incurring a reward or penalty, the awareness that one is being closely monitored: these features of performance management do more than just make agents accountable; they reshape agency itself.

Indeed, performance management is disciplinary, not just in the sense that it involves the allocation of penalties, but also in the deeper sense suggested by Michel Foucault: the use of organized techniques to produce self-regulating subjects (i.e., service providers) who, under conditions of apparent autonomy, conduct themselves in ways that are consonant with prevailing institutions, values, and interests.

~The Organization of Discipline, p2

PERFORMANCE OR PUNISHMENT?

The beginning of the twenty first century finds us in an era of governance by performance management and nowhere is this more the case than in welfare programs for the poor. By establishing outcome benchmarks focused on work participation and placement, federal and state officials define the goals of service provision and the terms of its evaluation. Choice sets for local actors consist primarily of alternative means toward mandated ends. With these benchmarks in place, sophisticated information systems are used to monitor frontline activities and measure priority outcomes. And based on the outcomes of performance competition, financial rewards and penalties are distributed to incent preferred behaviors and bring lagging service providers to heel. In welfare-to-work programs, performance systems now serve as the core technology for monitoring contract compliance; they guide decisions about when to renew or terminate contracts with local providers; they provide state officials with a yardstick and a prod for the achievement of program goals; and they constitute the major way in which state TANF programs are evaluated by federal officials.

~The Organization of Discipline, p2

PERFORMANCE OR PUNISHMENT?

Proponents rarely conceptualize performance management as a system of discipline. In celebratory rhetoric, it is presented as a way to harness the dynamic energies of markets, improve the evidentiary basis for policy choices, and reconcile policy experimentation with public accountability. The implicit promise is that local actors will be “freed” to go their own ways and then, later, will be judged by their performance and given the information they need to improve. The reality, however, entails a more complex interplay of structure and agency.

~The Organization of Discipline, p2