Introduction

Measurement & Evaluation of HCC Systems

Welcome!

Some took M&E I with Dr. Babu

I assume you learned about t-test, ANOVA, regression, logistic regression, mixed ANOVA, and linear mixed effects models

If you're a little rusty, please check out www.usabart.nl/eval!

Welcome!

Some took M&E I with me in **Spring 2017**

You finished >80% of Andy Field's book!

But note that I assume that you know this stuff at an "A level"

Feel free to revisit www.usabart.nl/eval

Welcome!

Some took M&E I with me in **Spring 2016**

You additionally learned about questionnaires CFA and SEM

The first part of this class will be repetition

- But we will go deeper at some points!

About the Class

Goals and Requirements

Goals of M&E I

M&E I taught you how to scientifically evaluate computing systems...

...using a quantitative, user-centric approach...

...with state-of-the-art statistical methods.

Final goal: being able to conduct experiments and evaluations yourself

New in M&E II

This advanced course will pay special attention to two things:

1. Subjective valuations of users‘ experience using multi-item psychometric instruments…

…and analyzing these scales using exploratory and confirmatory factor analysis (EFA and CFA).

New in M&E II

2. The evaluation of structured models of hypotheses using structural equation modeling (SEM).

We will also cover advanced methods such as Rasch modeling and factor mixture analysis (FMA).

Final goal: become a real stats wizard!

New in M&E II

After this course you will be able to write papers like these:

- http://www.academia.edu/download/46247454/viewcontent.pdf (scale development, EFA and CFA)
- http://bit.ly/recsys2012full (SEM, class example)
- http://130.18.86.27/faculty/warkentin/BIS9613papers/MISQ_SpecialIssue/BulgurcuCavusogluBenbasat2010_MISQ34_RationalityAwareness.pdf (scale development, EFA, CFA, SEM)
- http://escholarship.org/uc/item/1635g9cf.pdf (SEM with interaction effects)
- https://pure.tue.nl/ws/files/38719271/art_3A10.1007_2Fs11257_016_9178_6.pdf (multi-level / mixed SEM)
- http://bit.ly/privdim (EFA, CFA, FMA)
- https://1drv.ms/b/s!Ah3Xkc8v51g_nnvT7M_dBu-FMWS8 (model invariance)
- https://pure.tue.nl/ws/files/47009609/789228-1.pdf (Rasch modeling, p13-14)

About the Class

Everything can be found at usabart.nl/eval2

Hand in homework etc. via email

About the Class

Recap of methods and M&E I (week 1-3)

(Aug 28: Dr. Caine; Aug 30: online; Sep 4: McAdams 230)

Readings:

Knijnenburg and Willemsen handbook chapter, three chapters from Kline, a few slides from Muthen (optional)

This should be familiar…

If this stuff is not familiar, this is a good time to catch up!

About the Class

Path models (week 3-4)

“Mediation analysis on steroids”

Readings:

Kline chapters 6, 7, 11 and 12, some Muthen slides (optional)

This is the foundation for the more complicated stuff

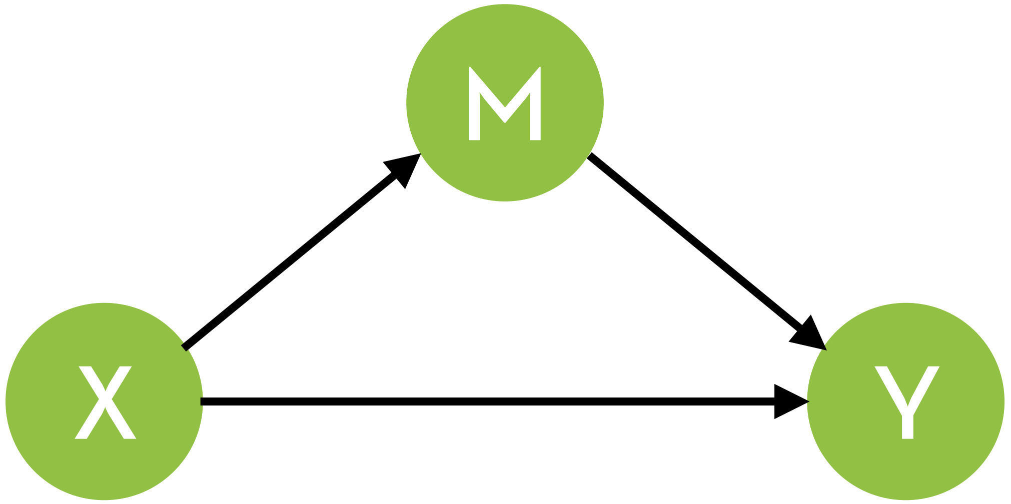
Make sure you don’t fall behind!

Path models

Mediation analysis:

X -> M -> Y

Does the system (X)
influence usability (Y)
via understandability (M)?



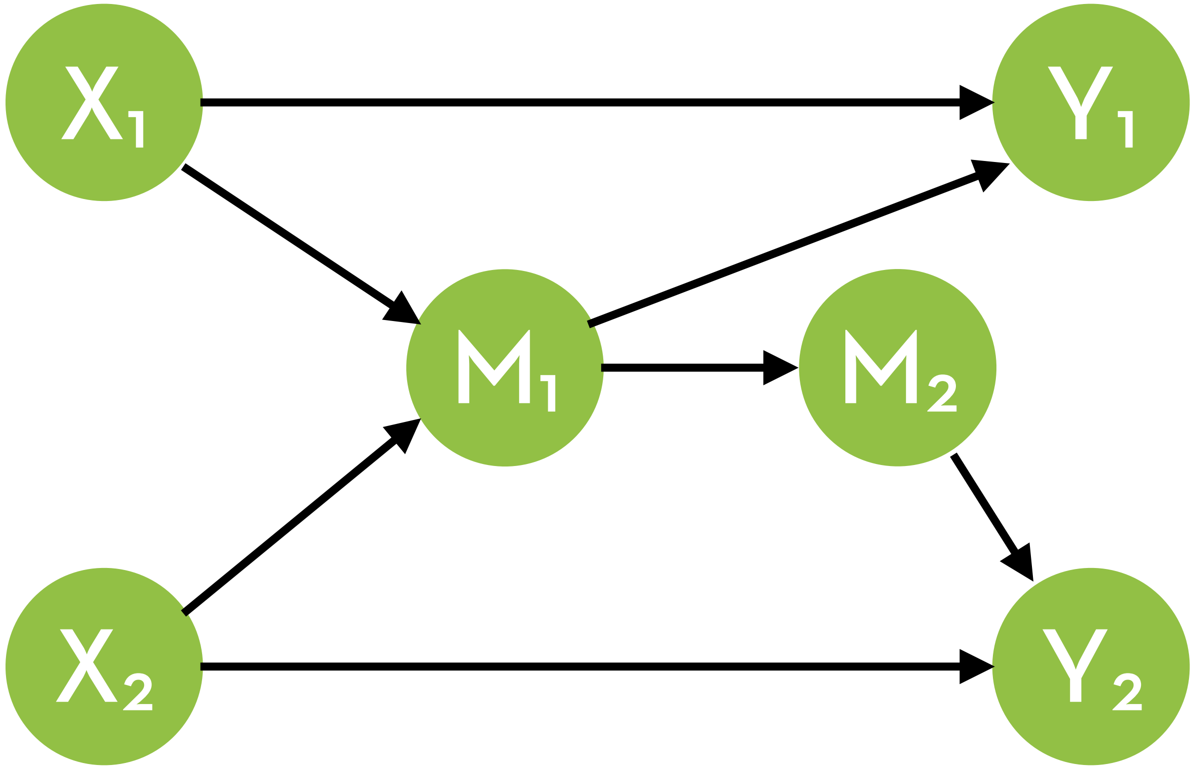
Path models

More complex models:

- What is the total effect of X1 on Y2?
- Is this effect significant?
- Is it fully or partially mediated by M1 and M2?

This is very tedious!

Path model analysis does this all at once



About the Class

Psychometrics (week 5)

“How to measure subjective valuations (such as satisfaction) with questionnaires”

Readings:

DeVellis chapters 1-5

This is more art than science

There is no single best solution when it comes to scale development

Psychometrics

Objective traits can usually be measured with a single question

- (e.g. age, income)

For subjective traits, single-item measurements lack **content validity**

- Each participant may interpret the item differently
- This reduces precision and conceptual clarity

Accurate measurement requires a **shared conceptual understanding** between all participants and researcher

Psychometrics

Solution: psychometrics

Use validated, multi-item measurement scales!

I will teach you how to:

- Find/use/adapt existing scales
- Develop new scales
- Use methods from usability testing to pre-validate your scales

About the Class

Factor analysis (week 7-8)

“How to analyze the quality of your questionnaires”

Readings:

Kline chapters 9 and 13, Loehin chapters 5-6, Tabachnick chapter 13, and a bunch of Muthen slides (optional)

We will cover both Confirmatory and Exploratory Factor Analysis

Note: EFA is the first topic that was not covered in the 2016 edition of M&E…

CFA

Is the scale really measuring a **single** thing?

- 5 items measure satisfaction, the other 5 convenience
- The items are not related enough to make a reliable scale

Are two scales really measuring **different** things?

- They are so closely related that they actually measure the same thing

We need to establish **construct validity**

This makes sure the scales are unidimensional

CFA

Factors are **latent constructs** that represent the trait or concept to be measured

- The latent construct cannot be measured directly

The latent construct "**causes**" users' answers to items

- Items are therefore also called **indicators**

CFA can establish whether a set of indicators exhibits convergent and discriminant validity

- It also turns the items into a normally distributed measurement scale

EFA

While CFA validates the fit of a factor, in EFA, the factor structure is "free"

- Effectively, it infers the structure from the data

Use EFA when you have no idea about the factor structure

- E.g. semi-related behaviors
- E.g. A (large) factor that didn't fit and might consist of multiple dimensions instead

About the Class

What about week 6?

Midterm on path models!

There are 3 midterms and a final (see later)

About the Class

Structural Equation Modeling (week 9-10)

"A combination of path models and CFA"

Readings:

Kline chapters 10 and 14, and a bunch of Muthen slides (optional)

I suspect that SEM will be your go-to evaluation method for outcomes of user experiments

This is where the 2016 version of M&E I pretty much ended…

About the Class

Structural Equation Modeling (week 9-10)

"A combination of path models and CFA"

Readings:

Kline chapters 10 and 14, and a bunch of Muthen slides (optional)

I suspect that SEM will be your go-to evaluation method for outcomes of user experiments

This is where the 2016 version of M&E I pretty much ended...

SEM

Combine **factor analysis** and **path models**

- Turn items into factors
- Test causal relations

Advantages:

- Powerful: factors are retained, you can compensate for measurement error
- Simple reporting: you get a path model that explains the effects

About the Class

Advanced SEM (week 11-13)

“For complex data and interaction hypotheses”

Readings:

Kline chapters 16-18, and a bunch of Muthen slides (mandatory!)

Multi-level SEM, measurement invariance, interaction effect

Some of this stuff is not covered in Kline; some of it is not possible in R!

Multi-level SEM

Repeated measurements

e.g. participants make 30 decisions

(Partially) within-subjects design

e.g. participants are randomly assigned to 1 of 3 games, and test it once with sound on and once with sound off

Grouped data

e.g. participants perform tasks in groups of 5

A combination of the above

Invariance

Are scales equally valid between groups?

- Can “satisfaction” be measured in the same way for older and younger users?
- Does the concept of “privacy concerns” mean the same thing between cultures?

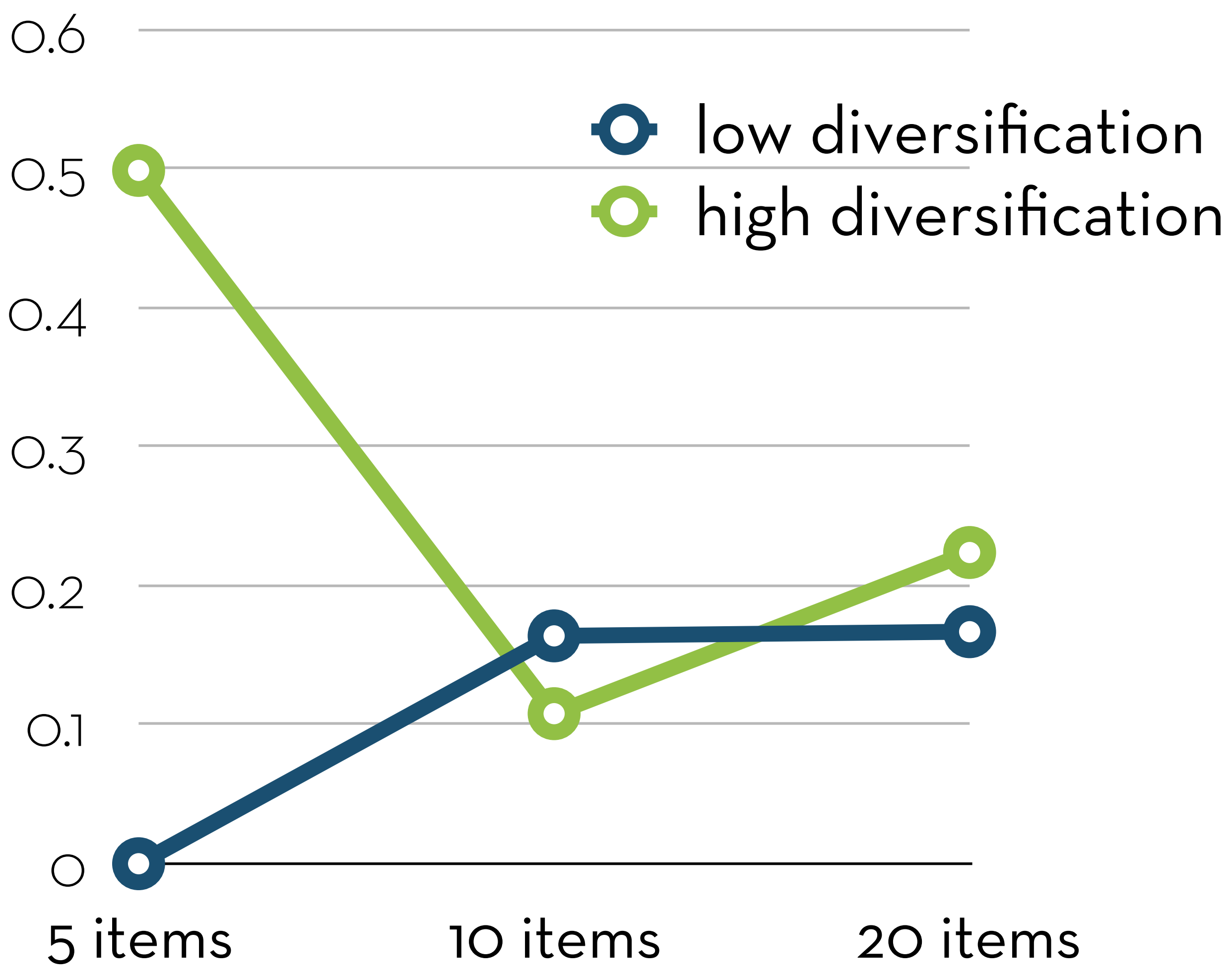
Measurement invariance testing can give you the answer to such questions

Interaction effects

Is the effect of diversification different per list length?

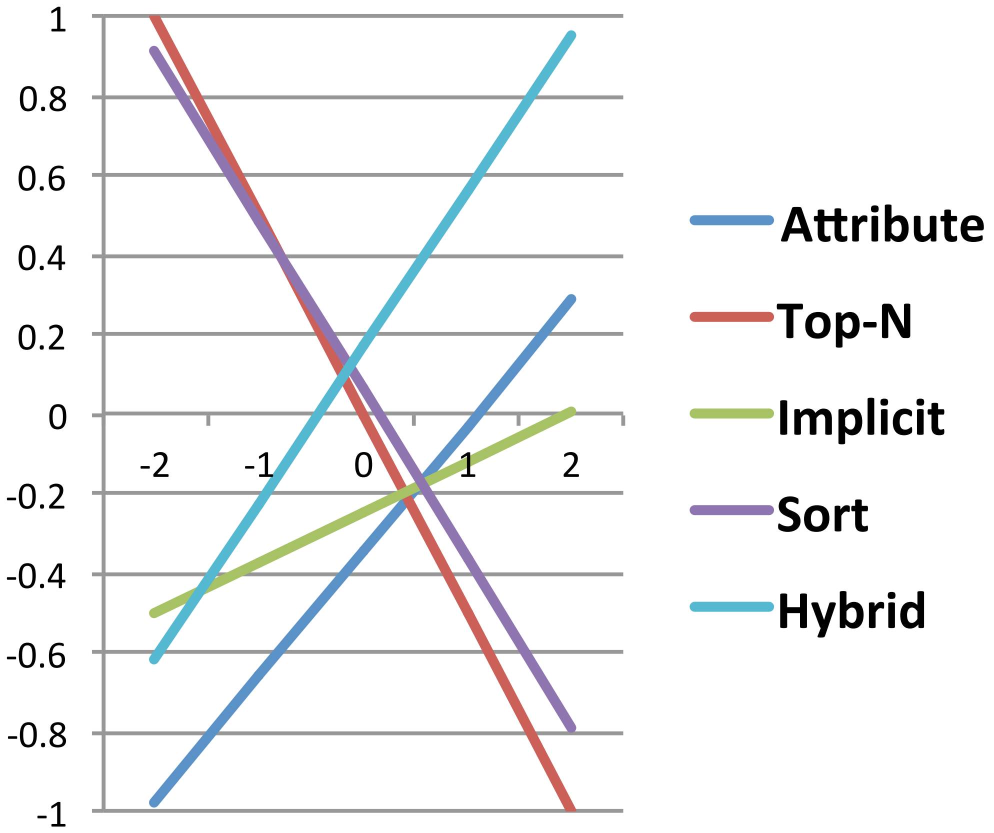
Is the effect of list length different for high and low diversification?

Perceived quality



Interaction effects

Does the effect of PE method depend on the user's level of domain knowledge?



Interaction effects

This is easy in regressions

Just multiply the dependent variables! y ~ x1*x2

In SEM, it depends on type of variables:

- manipulation * manipulation is easy (just create dummies)
- manipulation * factor is fussy, but doable (using multiple group models or predicted random slopes models)
- factor * factor is difficult (can only be done with predicted random slopes models)

About the Class

Advanced topics (week 15-16)

“Some cool stuff that’s a bit obscure”

Readings:

A bunch of Muthen slides (mandatory!), Bond chapters 3-4

Latent Categorical Analysis, Factor Mixture Analysis, Rasch Modeling

These methods are only used in special cases (but can be very powerful!)

LCA and FMA

Methods for **clustering** people into distinct groups…

…based on how they answer certain questions

…based on behavioral patterns

…etc

Two versions:

Based on "raw data": Latent Categorical Analysis

Based on factors: Factor Mixture Analysis

Similar to machine learning methods

Rasch modeling

Let’s say people answer a number of yes/no* questions

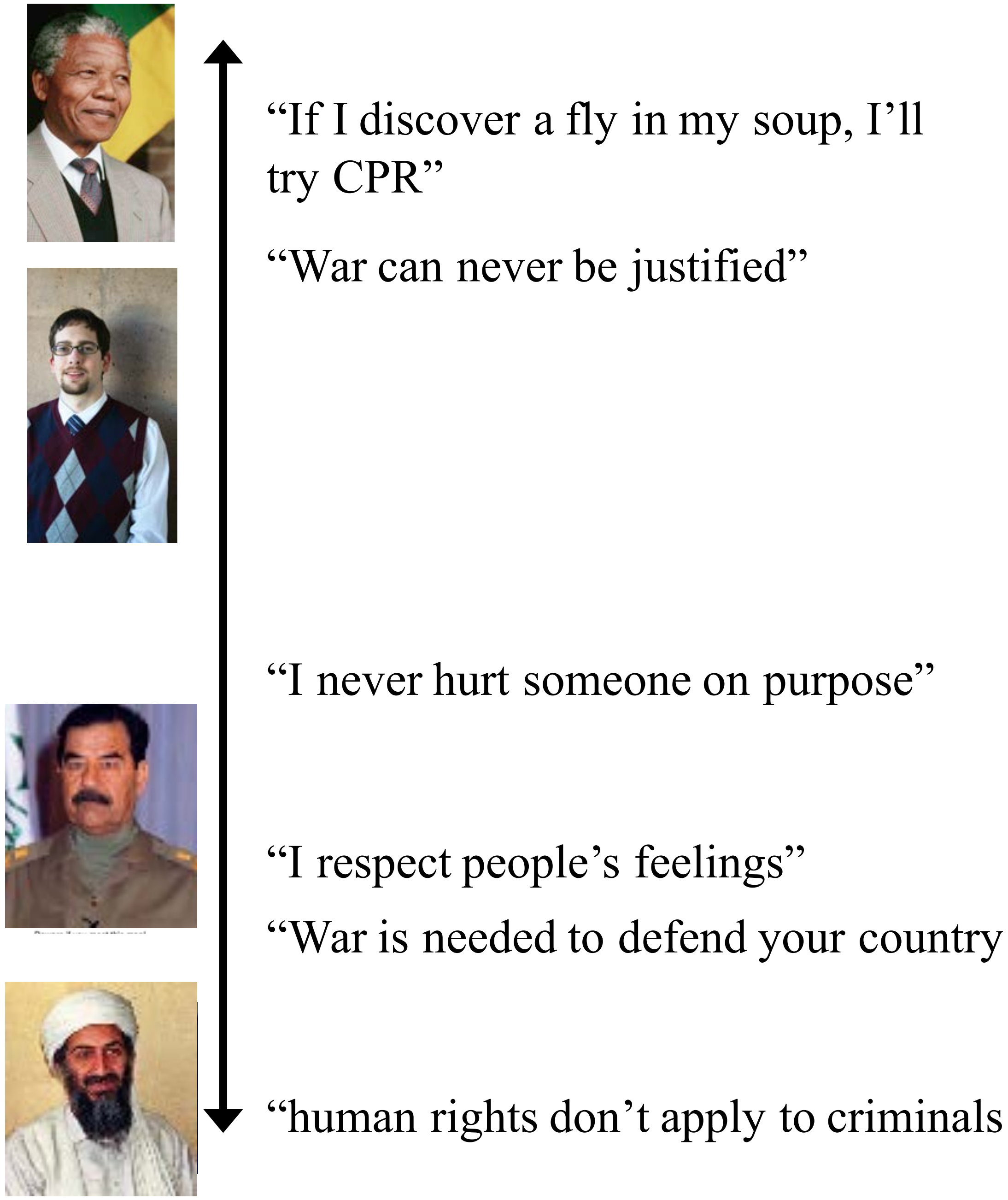
We can put both persons and items on a scale.

- The higher the score of a person, the more “able” this person is on this scale
- The higher the score of an item, the more “difficult this item is on this scale

Rasch modeling: establish this scale

- Benefit: can carefully inspect individual persons and items

Rasch modeling



“If I discover a fly in my soup, I’ll try CPR”

“War can never be justified”

“I never hurt someone on purpose”

“I respect people’s feelings”

“War is needed to defend your country

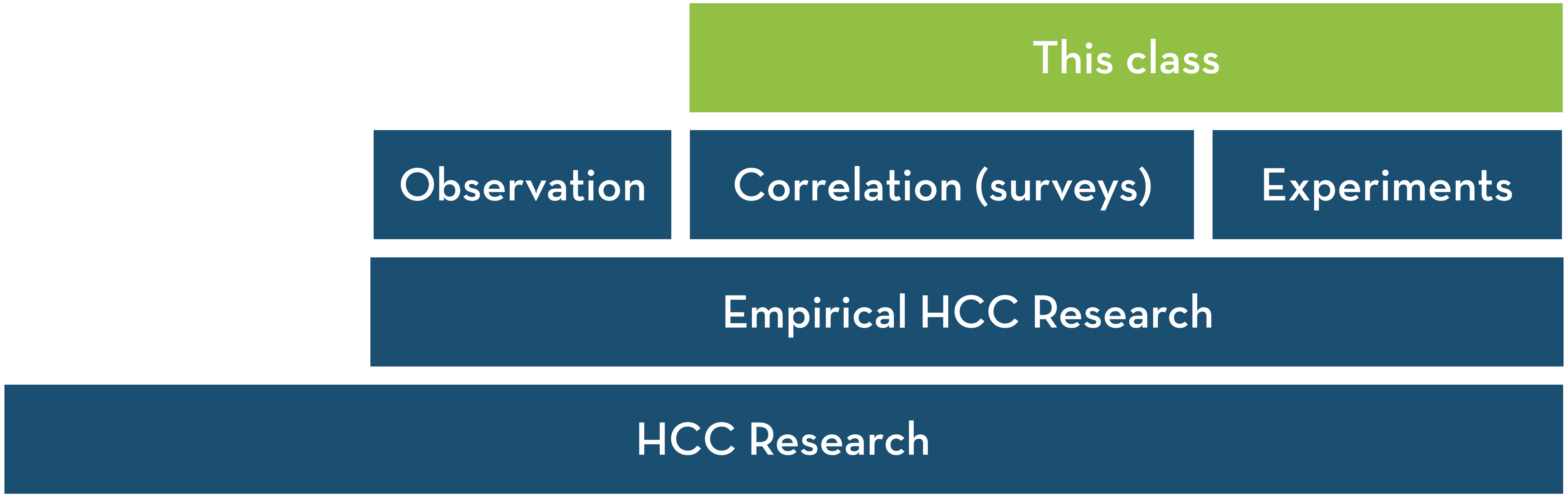
“human rights don’t apply to criminals

<– Example model

Rule: An item I and a person P have the same level if there’s a 50% chance that someone with the same trait level as P will comply with item I.

About the Class

Place within HCI:



This class

Observation

Correlation (surveys)

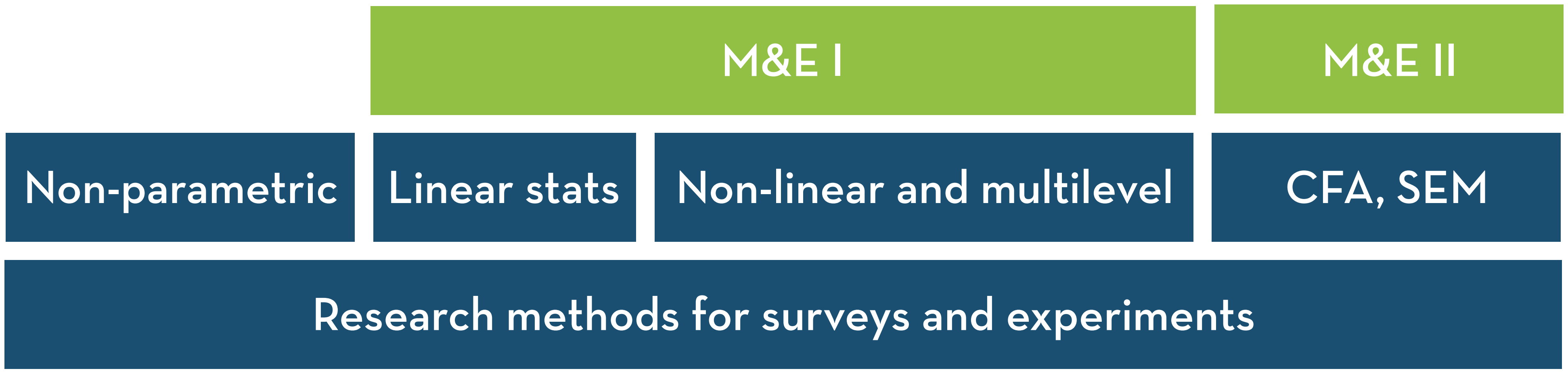
Experiments

Empirical HCC Research

HCC Research

About the Class

Place within HCI:



M&E I

M&E II

Non-parametric

Linear stats

Non-linear and multilevel

CFA, SEM

Research methods for surveys and experiments

What I Want From You

Requirements, Rules, Tips, Etc.

Software

We will mostly use R

- Please install R, RStudio, and the lavaan package

For some advanced methods, we will have to use MPlus instead

- It can also be used for the less advanced stuff (and is sometimes, but not always easier than R)
- Officially you will have to buy MPlus (full student edition)

I haven't yet decided on the software for Rasch modeling

Readings

Please read the assigned chapters **before** class

- I'm gonna assume you read them (lectures will be confusing if you didn't)
- You're gonna have to read them to do the homeworks anyway (better get it over with)

I know it is **a lot** to read each week

- The Muthen stuff is usually optional (more advanced), but explains more carefully how to do things in MPlus

Homeworks

4 homeworks (each 10%):

- HW 1: Path models
- HW 2: Psychometrics
- HW 3: EFA and CFA
- HW 4: SEM and multi-level SEM

Homeworks

Contents:

- Data analysis questions (should be done in R; provide code excerpts and explanation in your own words)

You are allowed to discuss the assignments, but you have to write your own write-up

- (i.e. you can discuss, but not copy)

HW2 is different, no data analysis

- Do it alone, please

Midterms + Final

3 midterms, 1 final (15% each)

- Midterm 1: Path models
- Midterm 2: EFA and CFA
- Midterm 3: SEM and multilevel SEM
- Final: Everything

Contents:

- Similar to assignments, but with a time limit
- Open book, open laptop (but no Wi-Fi)

Academic integrity

Be nice

Don’t cheat

Outside the class...

Feel free to come to me with statistics questions regarding your own research!

- I like doing this stuff
- I am usually consulting on ~10 different projects at a time