Lecture 5
Creating data by observing and measuring phenomena
DHX_MET1 Methodology 1

Stanislav Ježek

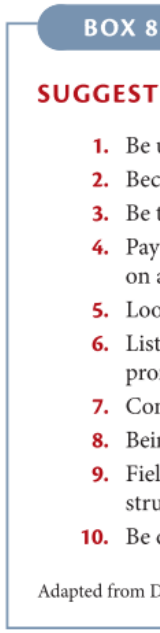
Faculty of Social Studies MU

- Observation concerns the planned watching, recording, analysis, and interpretation of behavior, actions, or events.

- (1) control (are the observations conducted in an artificial or in a natural setting?),
- (2) whether the observer is a member of the group that is observed or not (participant versus nonparticipant observation) – a dimension,
- (3) structure (to what extent the observation is focused, predetermined, systematic, and quantitative in nature), and
- (4) concealment of observation (are the members of the social group under study told that they are being studied or not?).

Participant observation

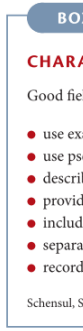
- the researcher gathers data by participating in the daily life of the group or organization under study.
- "to grasp the native's point of view, his relation to life, to realize his vision of his world"
- Somewhere between
  - Pure participation has been described as "going native"; the researcher becomes so involved with the group under study that eventually every objectivity and research interest is lost
  - Pure observation seeks to remove the researcher from the observed actions and behavior; the researcher is never directly involved in the actions and behavior of the group under study.
- Getting permission, access -- earning acceptance & rapport – choosing key informants
- Descriptive – focused – selective observation
- Writing field notes


BOX 8

SUGGEST

1. Be
2. Bec
3. Be
4. Pay on
5. Loc
6. List pro
7. Cor
8. Bei
9. Fiel stru
10. Be

Adapted from D


BO

CHARA

Good fie

- use ex
- use ps
- descri
- provid
- includ
- separa
- record

Schensul, S

In unstructured observation distinguish between the following

- „Pure“ descriptions of people‘s visible behavior
- Our interpretations of the observed behavior – emotions, intentions, motivations, meaning
- Our own feelings and impressions from the observations – the power of unconscious

Structured observation

- Observation of predetermined phenomena – a continuous dimension of the amount of sructure
- Observation of **behavior** (includes any observable characteristics, voluntary/involuntary, small like eyeblink to large like declaring war) of **one or more people** (interactions). Even oral/written language may be considered behahavior and observed. Also contextual/triggering events are included in observations.
- Ex. mystery shoppers, classroom observations, usability testing
- Observation schedule (coding scheme, sheet) - what
- Behavior sampling – when, where

Observation schedule

- Operationally defines **what** is to be observed
- Paper/electronic form for **recording** observations with some instructions, or recording software (e.g. http://www.boris.unito.it/)
- An observer must be **trained** in using a particular observation schedule
- Items
  - checks, checklist – record a behavior has been observed
  - rating scales – rate, evaluate, classify the behavior
  - the above can be put on a timeline(timescale), or electronically appended with timestamps

Observation schedule desiderata

- Clarity & Objectivity (after training) – minimal subjectivity left in what is and is not behavior of interest and how to record it
- Mutually exclusive and collectively exhaustive categories
  - Observed instance of behahavior should fall in one category only
  - Every relevant instance of behavior should fall in some category
- Minimal cognitive burden, ease of use
  - Choices whether a behavior is relevant and how to code (classify, rate) it should be easy, automatic after training
  - More difficult coding is to be done in the analytical phase, or from video-recordings
- Minimal use of observer‘s memory
- In people, focus on observable behavior, not on inferring emotions, intentions, motivations….

- Inter-observer(coder) agreement as a check of the above
  - *Kappa* coefficients – measure agreement beyond chance, (-1)…0…1

Sources of Observer bias and errors

- Insufficient training
- Reactivity to being observed
- Selective attention and attention fluctuation
- Memory errors – forgetting, distortions (schematizations)
- Social cognition distortions – e.g. halo effect and other first impressions, projections
- Fatigue
- Observer drift – systematic individual changes in applying the observation schedule – the need for re-training

Behavior sampling
When to observe

- Just like we cannot observe all people we usually cannot observe all their behaviors
- Time sampling
  - select time intervals & durations systematically or randomly
  - good use of available time switching between focused attention and rest
  - usually for frequently occurring behavior
  - e.g. all personell-customer interactions are recorded for 1 min every 15 minutes in a restaurant
- Event sampling
  - sample of events, usually systematic
  - all relevant behavior is recorded for the whole selected event
  - e.g. every 10th guest‘s interactions with the personnel is recorded

Using measurement instruments to capture physiological correlates of psychological experience

Activation – sympathetic nervous systém, quick acting

- Cardioascular indicators – blood pressure, heart rate, breathing depth and frequency
- Skin conductivity – galvanic skin response
- Muscle tone, incl. face muscles - electromyography

Attention – eye-tracking

Advantages and disadvantages of observation

- Directness
- Control over reliability and validity of observation
- Richness of data when recording is used

- Takes a lot of time and other resourses
- Requires access to behavior (or ability to manipulate)
- Behaviour may be affected by observation

CONSTRUCTION OF MEASUREMENT INSTRUMENTS - SCALES

- What do we wish to measure – CONSTUCTS
- CONSTRUCTS are concepts which are part of our theory explaining observable behaviour of people (or groups, organizations)
- e.g. need for cognition, cohesion, climate
-
- ~~Measurement is the assignment of numbers or other symbols to characteristics (or attributes) of objects according to a prespecified set of rules.~~ Traditional positivistic, but useless S.S. Stevens' definition of measurement.
- In practice, measurement – scaling – is an attempt to
  - a) create a scale reflecting the quantitative property of a construct and
  - b) create an instrument allowing us to map measured objects on this scale

QUANTITATIVE PROPERTY OF A CONSTRUCT – based on theory

- **Nominal** construct
  - characteristic that has values that differ in quality, not in quantity
  - e.g. sex of a person, preferred color, occupation
  - object may only be the same or different in this characteristic
- **Ordinal** construct
  - has values that differ in quantity enough to allow us to order them
  - often the values are categories
  - e.g. education level, grade in school
  - objects may be ordered acc. to this characteristic (transitivity)
- Quantitative construct
  - has values that may be ordered and the distance between values may be defined – numeric scale, discrete/continuous makes sense
  - a distance of one can be (carefully) thought of as a unit
  - **Interval** construct - distance between objects is defined (deg Celsius, IQ, risk av.)
  - **Ratio** construct – has absolute 0, ratios of objects are defined (deg Kelvin, ?)

Relationship between Quantitative property of a construct and a **SCALE** we want to build

- Scales are also nominal, ordinal, interval and ratio – **scale level**
- **Scale level** defines meaninful mathematical relations/operations on the values
  - **=, not =**
  - **=, not =, <, >**
  - **=, not =, <, >, +, -**
  - **=, not =, <, >, +, -, x, /**
- For a construct we can build a scale on the same level or **lower**

- Many combinations are possible.

X

- From the perspective of operationalism, the scale defines the construct

OPERATIONALIZATION OF A CONSTRUCT

- Deriving, from theory, all possible **manifestations** of the construct we are trying to measure
- There are many, so some map/tree of them is usually necessary
- Major areas, sets of manifestations of the same kind are often termed **FACETS** (S-B call them *dimensions*, which is a bit misleading)
- Based on theory, how should people(groups, organizations) with different levels of the construct…
  - …currently behave, act (in various situations, even in responding to questions)
  - … feel
  - … be perceived by others
  - … have achieved in the past
  - … should be their future goals……
- Then we think of ways to observe/capture as many manifestations
  - unobserved manifestations (for whatever reasons) limit validity (coverage)

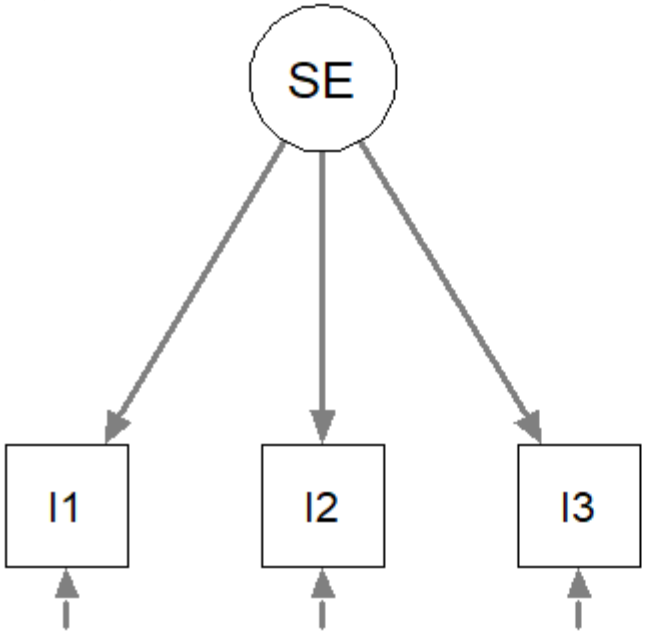
SCALING

- Observed manifestations – **indicators, items**
- **Scale** is built from **items** by various statistical scaling techniques.
- Currently most used is **Likert scaling** technique using **Likert-type items**.
- The scale value (interval) is created by summing/averaging item responses (ordinal).
- This simple procedure is based on a large number of assumptions together forming an underlying reflective measurement model.

Models of Measurement

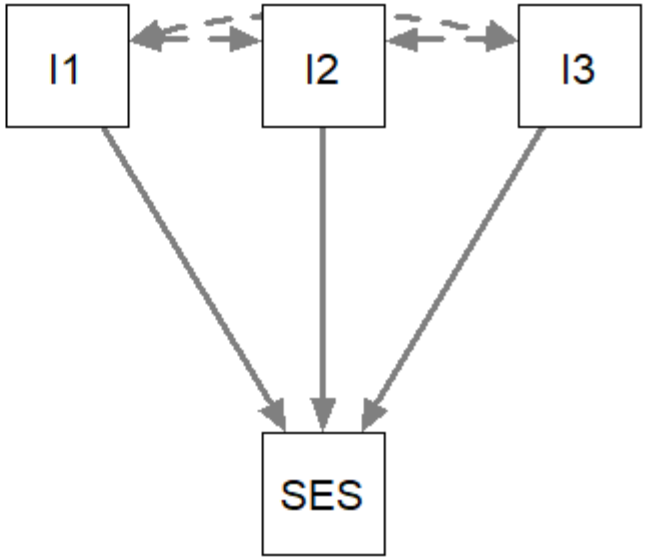
Reflective model
(e.g. self-esteem)

- I1: On the whole, I am satisfied with myself.
- I2: I feel that I have a number of good qualities.
- I3: I am able to do things as well as most other people.



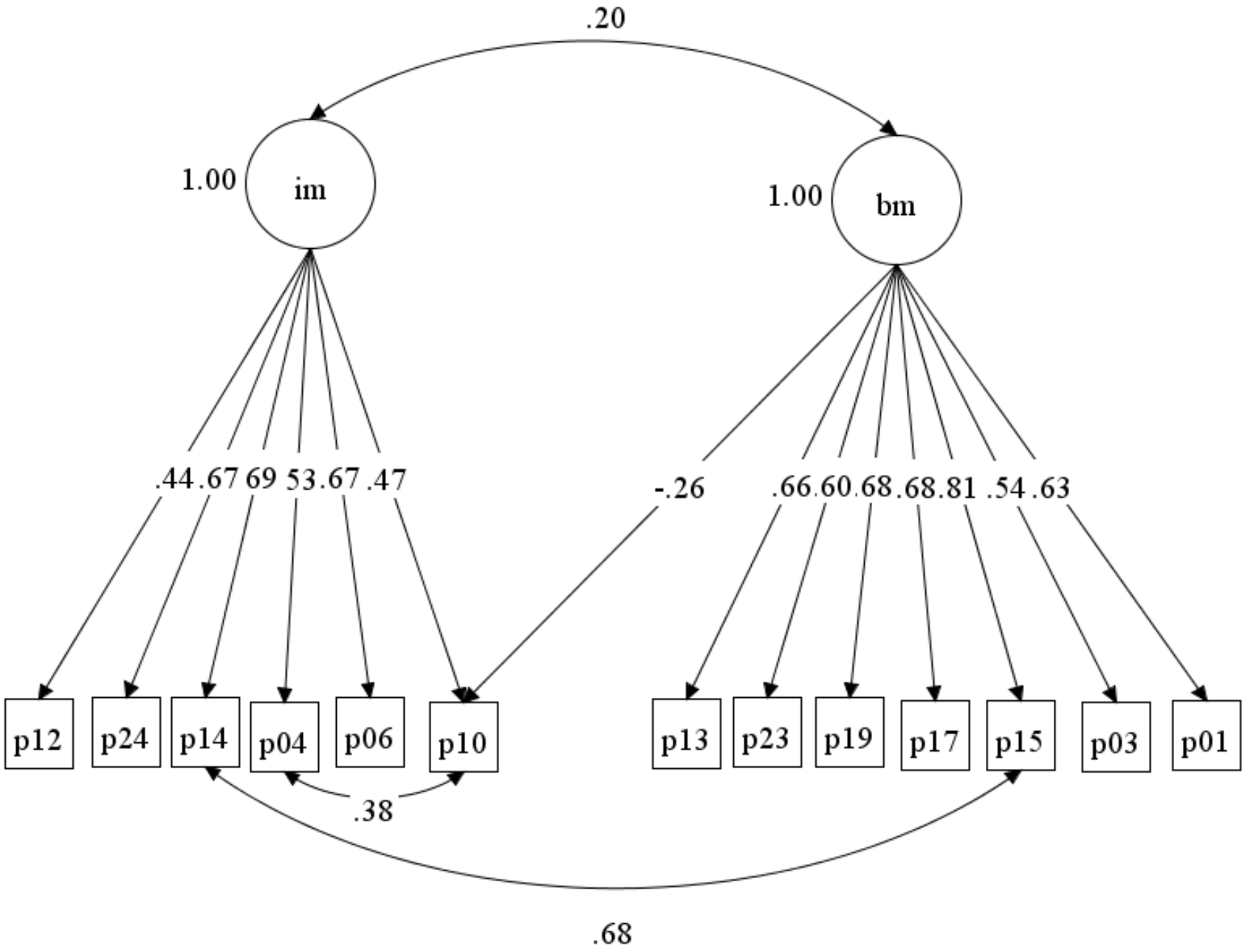
Formative model
(e.g. socio-economic status)

- I1: Respondents education.
- I2: Parents education.
- I3: Income level.



The assumptions in the reflective model

- The construct is a latent continuous quantitative variable (may also be nominal - latent class models)
- There is only one construct (latent variable, factor)
  - If there are more they must be added to the model
- Item responses are only due to construct(s) and random error – residual variance
  - Causality is explicit here
- Items correlate only due being caused by the same construct
  - Local independence of items


.20
1.00
im
1.00
bm
.44
.67
.69
.53
.67
.47
-.26
.66
.60
.68
.68
.81
.54
.63
p12
p24
p14
p04
p06
p10
p13
p23
p19
p17
p15
p03
p01
.38
.68

- **Reliabilit**y is defined through this model
  - The proportion of residual error variance across the items
  - McDonald‘s omega

- Cross-loadings and correlations between residuals increase construct-irrelevant variance in a summation score – i.e. affect validity

- Sometimes model is simplified by excluding „bad behaving“ items – may negatively affect validity

- If a model is more complicated than assumed
  - latent variable should be used in further analysis instead of summation score
  - measure should be improved

- Development of a measure im practice = looking for items functioning according to the assumptions while covering as many facets as possible.

DIRECT SCALING – RATING/RANKING SCALES

- Often, instead of sophisticated measurement, we assign the scale values directly – ourselves or ask respondents
- In this mode a single rating item is called scale and represents a single construct
  - sometimes the construct is implicit, which limits critical interpretation
  - confusion is created about what is meant by a scale
- The measurement model is in the head of the rater
  - If it were a part of observation study he/she would be trained….
  - All imaginable sources of bias need to be considered
- Validity & reliability of single rating/ranking items is determined mainly by comparing to criteria and repeated administration.
  - Unless you have some empirical data in favor of validity & reliability of sinle items, be very, very suspicious

SUMMARY

- Prefer used well validated measures

- Observe as much as your resources allow you
- Interview as much as your resources allow you after you have observed at least a bit
- Survey only after you have gained knowledge – theory, observation, interview
- Be sure to distinguish between survey items asking for direct scaling and items of measurement scales

- Elaborate your knowledge of what is meant by validity and reliability