

MUNI

Faculty of Education



EERA Summer School 2019

Questionnaire adaptation

Jan Mareš, Kateřina Vlčková, Petr Květon

Příspěvek vznikl v rámci metodologických cílů projektu GAČR Strategie řízení třídy u studentů učitelství a zkušených učitelů (jejich “cvičných učitelů”) na druhém stupni základní školy (GA16-02177S).

Is it really simple to conduct a survey using questionnaires?

- Many novice researchers conduct survey using questionnaires
 - Because these measures appears simple to construct and administer
- But the collection of evidence to demonstrate the measure's validity and reliability requires expertise and resources
 - => whenever possible select a well-developer questionnaire => save the time required to develop your own measure

Sharing of experiences

Share in groups of 6
(6 minutes activity)

- Do you have experience with questionnaire adoption/adaptation/development?
- What was the questionnaire measuring (topic)?
- What were the main challenges you encountered? How did you solve them?
- What are according to your experience the main rules for adopting/adapting/developing an questionnaire?

Summarizing experiences with questionnaire use

Why to use „adopted“ questionnaires?

- Developing your own is a hard job

See EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests v. 4.2.6



- English language as the science „lingua franca“
 - It means better chance for publication of your results in the future



known scientists

vs.



local scientists

How to work with a renowned questionnaire?

Traditional perspective (1)

- Take it as it is from any available source (instructions, norms, questions)...and ask some students to do translation 😊
- It often does not work

Traditional perspective (2)

- Take it (a) from any available source and/or (b) write an email to author and ignore any answer including © word;
- Translation, back translation and if it looks similar, collect data and proceed EFA on them
- Better, but it often does not work either

How to work with a renowned questionnaire?

Currently there are some rules (ITC) which you should to follow.



Why so many rules?

Different traditions
in educational
science

Sociological:

one good question is enough

Psychological:

there are many sources of
variance, better to use scale
(several questions)

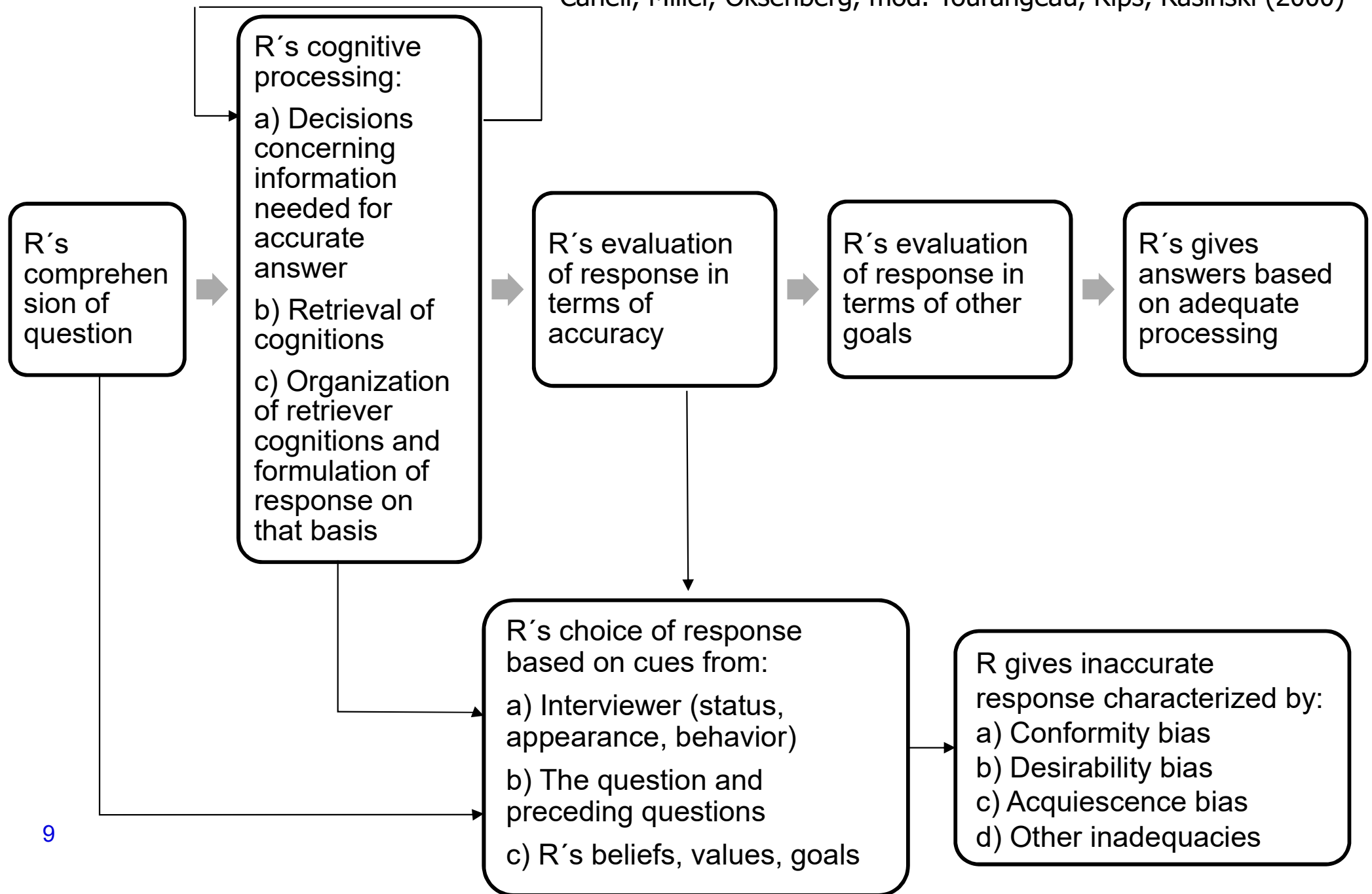
....

Pedagogical:

we have to adopt this
construct and _ they _ do it
this way

Model of the survey response process

Canell, Miller, Oksenberg, mod. Tourangeau, Rips, Rasinski (2000)



MUNI

And what are the issues?

Culture

Language

Methodology (research context)

Statistics

Culture and theory

Education is a part of the culture

- There are some differences in context, meanings and perspectives.
- The topic of school climate can serve as an example.
- *Q: Do you feel safe at school?*



Culture and theory

– *Q: Have you ever brought a gun to school?*

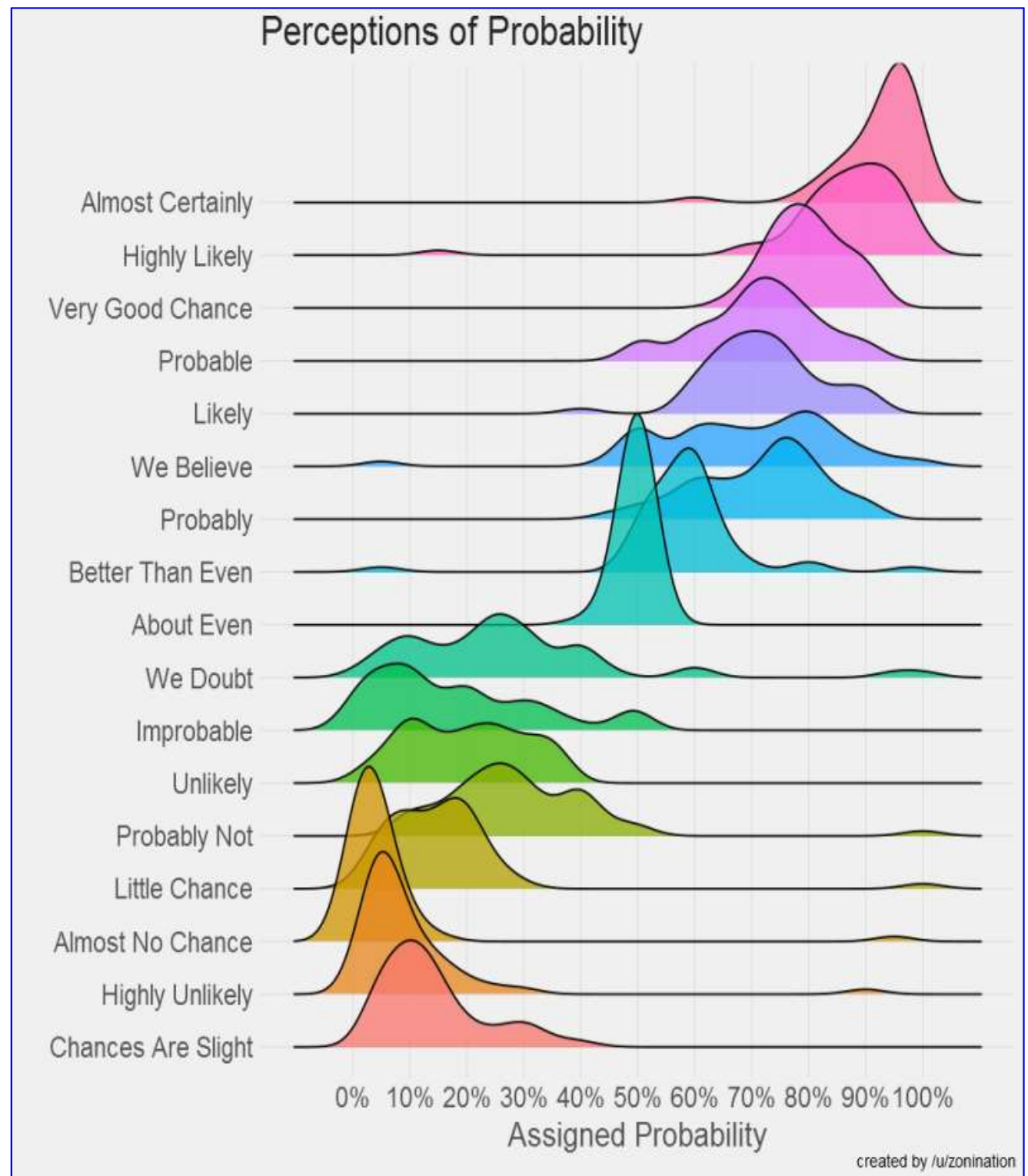


Language

There are differences in tone etc.

even before translation...

<https://blog.revolutionanalytics.com/2017/08/probably-more-probably-than-probable.html?fbclid=IwAR0ACCR5Hfpobj4XWGS1wjusbUG-wilyk60gOJVR3SmhANWAtwFVjeRd8>



Language

What does it mean?

„*in school*“

„*extra lessons*“

Bray, M., & Kobakhidze, M. N. (2014).
Measurement issues in research on shadow
education: Challenges and pitfalls encountered
in TIMSS and PISA. *Comparative Education
Review*, 58(4), 590-620.

Or in Likert scale:

How often x often 😊

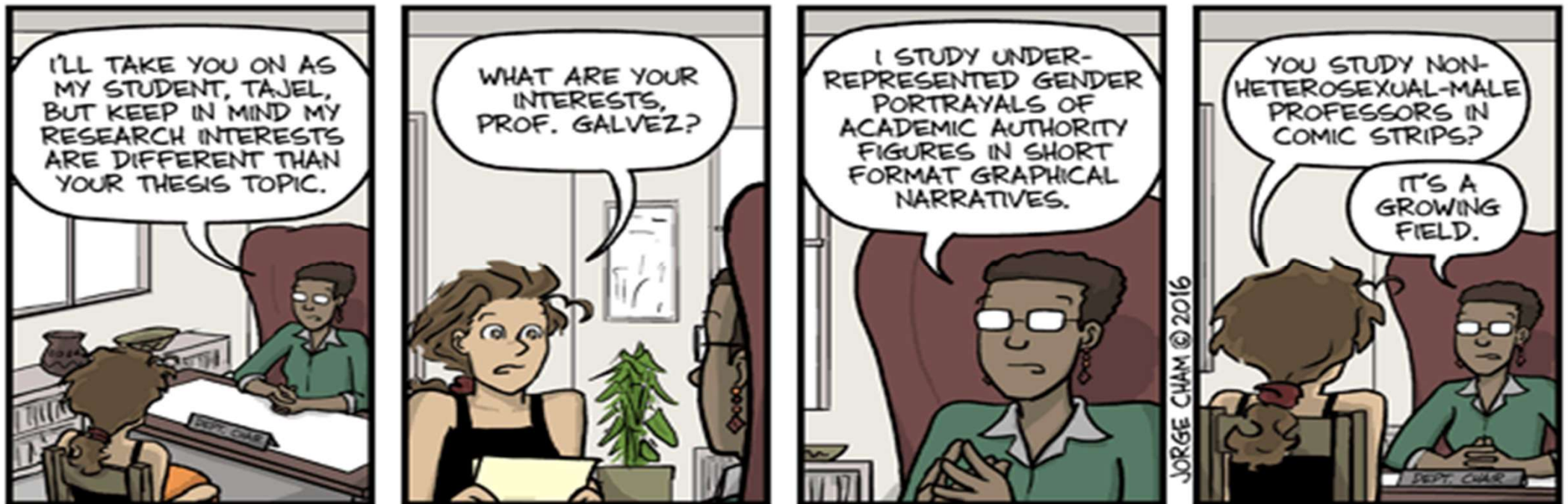
Methodology

– Typically context of other methods in the research
(how many items could respondents survive)

– Availability of the sample

Hanel, P. H., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PloS one*, 11(12), e0168354.

– Pen and pencil or online survey



MUNI



ITC GUIDELINES FOR TRANSLATING AND ADAPTING TESTS

The International Test Commission (ITC) is an “Association of national psychological associations, test commissions, publishers and other organizations committed to promoting effective testing and assessment policies and to the proper development, evaluation and uses of educational and psychological instruments.” (ITC Directory, 2001).

ITC GUIDELINES FOR TRANSLATING AND ADAPTING TESTS (1)

Pre-condition guidelines

PC-1 (1) Obtain the necessary **permissions** from the holder of the intellectual property rights relating to the test before carrying out any adaptation.

PC-2 (2) Evaluate that the amount of **overlap in the definition** and content of the construct measured (by the test) in the populations of interest is sufficient for the intended use (or uses) of the scores.

PC-3 (3) Minimize the influence of any **cultural and linguistic differences** that are irrelevant to the intended uses of the test in the populations of interest.

Test development guidelines

TD-1 (4) Ensure that the adaptation process considers **linguistic, psychological, and cultural differences** in the intended populations through the choice of **experts** with relevant expertise.

TD-2 (5) Use appropriate **translation** designs and procedures to maximize the suitability of the test adaptation in the intended populations.

TD-3 (6) Provide evidence that the test instructions and item content have similar **meaning** for all intended populations.

TD-4 (7) Provide evidence that the item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures are **suitable for all intended populations**.

TD-5 (8) Collect **pilot data** on the adapted test to enable item analysis, reliability assessment and other small-scale validity studies, so that any necessary revisions to the adapted test can be made.

Confirmation guidelines

C-1 (9) Select **sample** with characteristics that are relevant for the intended use of the test and of sufficient size and relevance for the empirical analyses.

C-2 (10) Provide relevant statistical evidence about the **construct equivalence, method equivalence, and item equivalence** for all intended populations.

C-3 (11) Provide evidence supporting the **norms, reliability and validity** of the adapted version of the test in the intended populations.

C-4 (12) Use an appropriate **equating design** and data analysis procedures when linking score scales from different language versions of a test.

Administration guidelines

A-1 (13) Prepare **administration materials** and instructions to minimize any culture- and language-related problems that are caused by administration procedures and response modes that can affect the validity of the inferences drawn from the scores.

A-2 (14) Specify **testing conditions** that should be followed closely in all populations of interest.

Score scales and interpretation guidelines

SSI-1 (15) Interpret any **group score differences** with reference to all relevant available information.

SSI-2 (16) Only compare scores across populations when the level of **invariance** has been established on the scale on which scores are reported.

Documentation guidelines

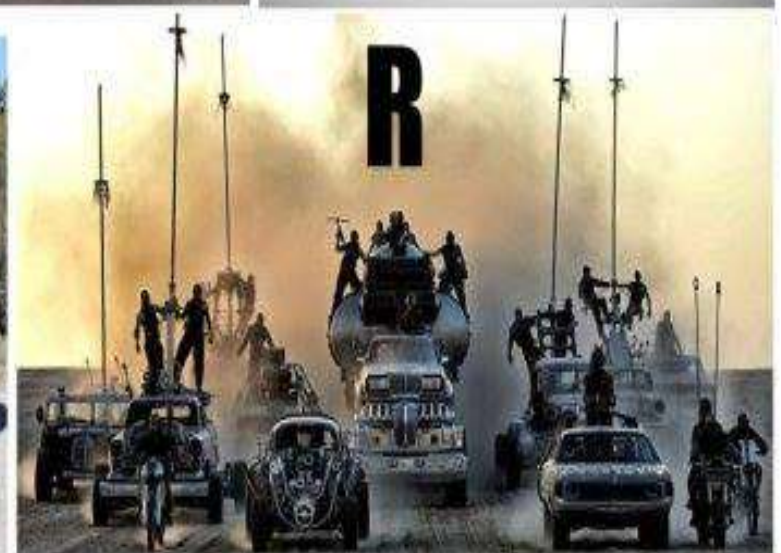
Doc-1 (17) Provide technical **documentation of any changes**, including an account of the evidence obtained to support equivalence, when a test is adapted for use in another population.

Doc-2 (18) Provide **documentation for test users** that will support good practice in the use of an adapted test with people in the context of the new population.

M U N I

Statistic issues

If statistics programs/languages were cars...



The process of assessing construct validity

Key moments

Theoretical evaluation of measured construct

- Background theory
- Dimensionality of the construct
- Face validity of individual items (cultural bias)

Empirical evaluation of the instrument

- Planning phase
 - Design (test-retest, cross-sectional)
 - Sample
 - Selection process (target population)
 - Sample size
- Evaluation of factor structure
- Network of relationships
 - Criterion validity (relations with instruments measuring the same construct)
 - Construct validity (relations with other variables according to theoretical expectations)
- Evaluation of reliability

Data collection

What is the minimum acceptable sample size? (with CFA in mind)

- Depends on many factors.
- Typical min = 5 respondents/parameter estimated, 10/parameter preferred (Bentler, 1995).
- When not MV normal – 15 respondents per parameter (Stevens, 1996)
- Typical case scenario $N = 200$ (Kline, 1998b; Loehlin, 1998; Boomsma and Hoogland, 2001).

Equivalency of measurement: statistical approach

- Equivalency of structure
- Equivalency of measurement
- Equivalency of true scores

Before statistical analyses

- Theoretical evaluation of measured construct
 - Background theory
 - Dimensionality of the construct
 - Face validity of individual items (cultural bias)
- Planning of empirical evaluation
 - Design (test-retest, cross-sectional)
 - Sample
- Supporting evidence
 - Criterion validity
 - Construct validity

Evaluating factor structure

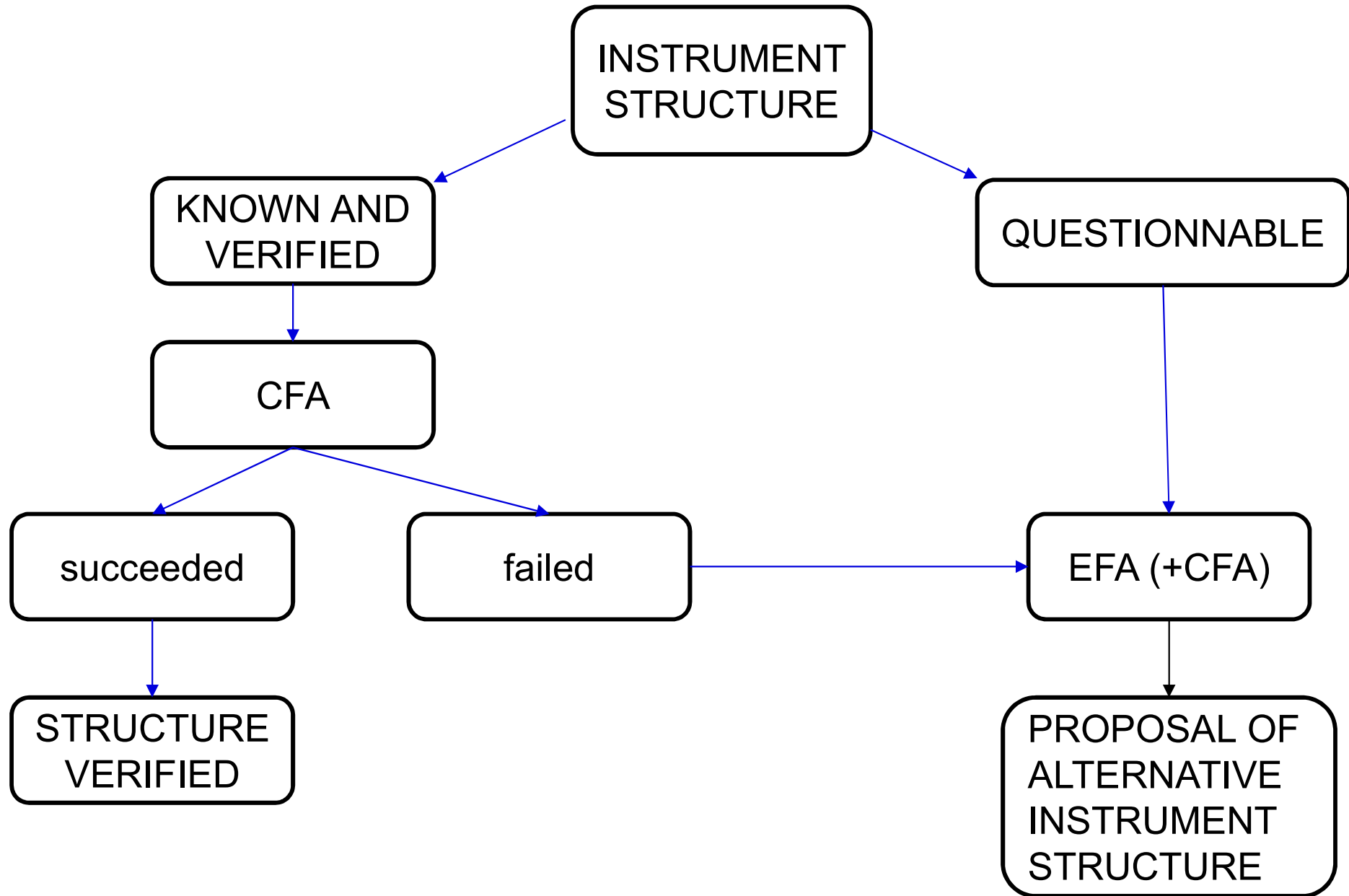
EFA or CFA or both?

- Inconsistent recommendations, different opinions

Typical scenarios

- Newly developed instrument
- Existing instrument with known (stable) psychometric characteristics and dimensionality
- Existing instrument with questionable dimensionality

EFA or CFA or both?



Performing CFA

- Evaluating known factor structure

Key decisions before

- **Sample**
 - Population
 - Sample size
- **To use Pearson or polychoric correlation matrix as the input?**
 - Serious issue
 - Earlier studies commonly used Pearson correlation, but violated assumption of continuous character of the data.
 - Likert items are ordinal in nature, so rigorously polychoric correlations are the correct option.
 - Polychoric correlations – sometimes computational problems (-- more parameters used -- bigger sample size needed)

Performing CFA: Does the model fit our data?

Fit indices

- Absolute Fit Indices
 - Chi-square, Relative chi-square (χ^2/df): should be less than 5 (or 3 when more strict).
 - Hoelter's CN, **SRMR**, GFI, AGFI, AIC, BIC
- Relative Fit Indices
TLI, **IFI**, NFI
- Parsimonious Fit Indices
PGFI, PNFI, PNFI2, PCFI
- Noncentrality-based indices
RMSEA, **CFI**, RNI, CI

Be sure to
report more
than one
index

Performing CFA: Does the model fit our data?

Decision about model fit

- Always check individual parameters (i.g. factor loadings of items), not only overall indices.
- Mostly not strictly black or white.

Three types of model fit

1. Apparently OK
(all indices below or above thresholds)
2. Apparent misfit
(all indices noticeably bad values)
3. Grey zone
some indices are OK, some are not
unfortunately the most frequent real scenario

Most frequent reasons for computational problems

- „overcorrelated“ items or factors

CFA failed to confirm factor structure?

Less severe misfit can be solved using modifications in the model.

- Removing items with low factor loadings.
- Releasing model constraints
 - Typically uncorrelated factors
- Adding correlation between errors
 - Needs to be theoretically justified
 - Modification indices can be helpful but can't justify by itself

Still not OK, what can I do?

- EFA seems to be the only way out.

Performing EFA

KMO interpretations

0.00 to 0.49 unacceptable

0.50 to 0.59 miserable

0.60 to 0.69 mediocre

0.70 to 0.79 middling

0.80 to 0.89 meritorious

0.90 to 1.00 marvelous

KMO Measure of Sample Adequacy

Method of extraction

- Principal component vs. “real” EFA
- When the goal is to identify latent factors, EFA is the way
Mostly, EFA methods produces similar results

How many factors to keep?

- Index with the longest history – Kaiser criterion and Cattell’s scree test (visual inspection)
- Use a set of methods like Velicer’s MAP, Parallel analysis, Optimal coordinates, or other
- What if indices are inconsistent?

How to rotate the factor solution?

- Oblique rotation (e.g. Direct Oblimime) is the best choice because in social sciences constructs are usually correlated
- Orthogonal rotation (mostly Varimax) is mostly used because of its simplicity but is recognized overly restrictive (factors are considered independent)

Number Factors: CD

Comparison Data

 Analysis Perform analysis only till nonsignificance improvement

Missing Data

 Delete cases listwise before doing the analysis Keep missing data

Correlations employed:

Spearman

Largest number of factors (must be less than the number of variables selected):

8

Size of finite populations of comparison data:

10000

Number of samples drawn from each population:

500

Alpha level when testing significance of improvement adding factor:

0.3

Continue

Cancel

Help

This dialog requires the R plugin and the following R packages: polycor, psych, GPArotation, nFactors, corpcor, ICS, R.utils.

OK

Paste

Reset

Cancel

Help

Number Factors: MAP-VSS-PA-OC-AF

Correlation Matrix to Analyse:

Heterogenous (Two Step)

Velicer's MAP

 Analysis

Very Simple Structure

 VSS Fit the diagonal

Number of factors to extract (more than hypothesized):

5

Extraction:

Maximum Likelihood

Rotation:

Obimin

Parallel Analysis and Non Graphical Scree Test

 Analysis

Number of samples:

1000

Quantile of the simulations:

0.50

Simulations based on:

 Mean Quartile

Model

 Components Factors

Samples from:

 Normal distribution Permutation

If "Permutation" selected:

 Delete cases listwise before doing permutations Do permutations with replacement

Correlation matrix for data permutations:

Pearson

Continue

Cancel

Help

This dialog requires the R plugin and the following R packages: polycor, psych, GPArotation, nFactors, corpcor, ICS, R.utils.

OK

Paste

Reset

Cancel

Help

Suggested alternative instrument structure

- In case of relatively large sample size you can perform EFA on randomly selected part of sample and confirm its results on the rest of the sample using CFA.

Example 1:

Video game addiction scale

- Evaluation of factor structure of the Video game addiction scale

Instrument

- 21 items
- 7 dimensions (3 items per dimension)
- Response format: 5point Likert scale (*never to very often*)

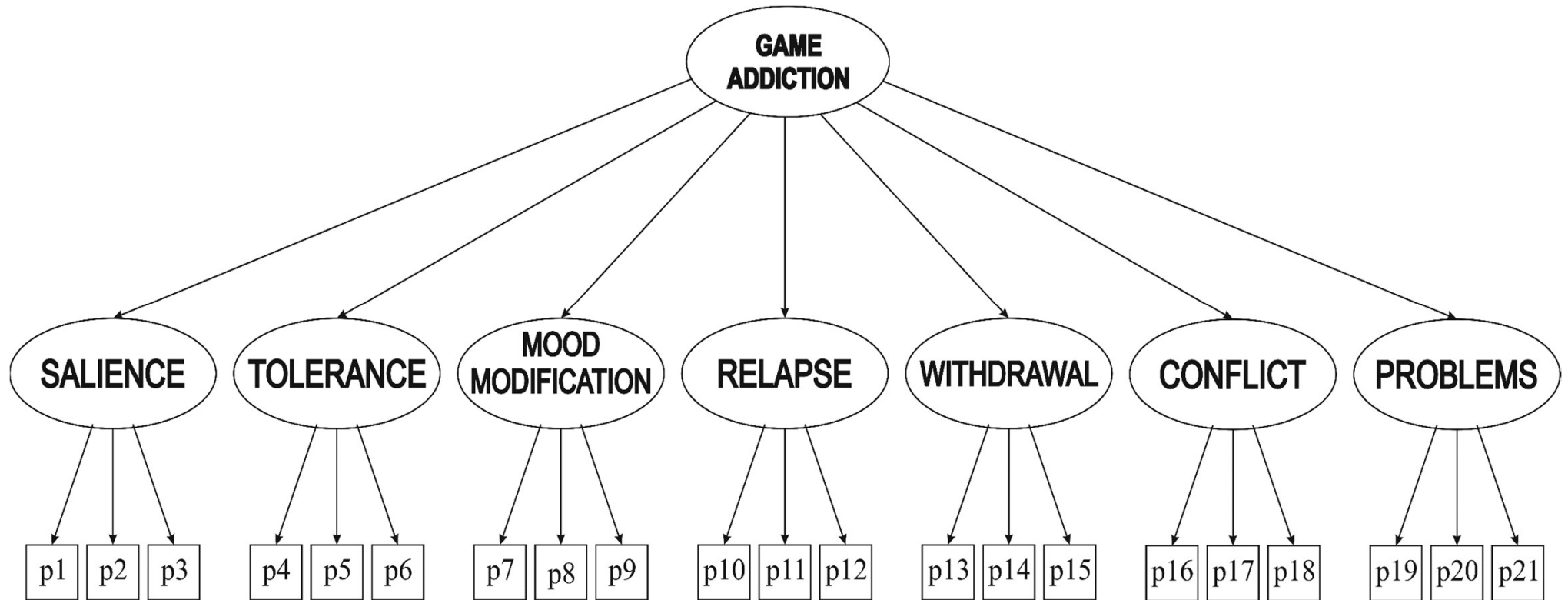
Research aim

- Evaluation of the originally proposed hierarchical structure with general addiction (first-order factor) and 7 dimensions (second-order factors)

Sample size

- $N = 423$ (approximately 10 vectors per parameter)
- 49 free parameters
 - 7 first-order factor loadings, 21 second-order factor loadings, 21 error terms

Example 1: Video game addiction scale



Example 1: Video game addiction scale

```
VGAmode1 <- `
SAL =~ p1 + p2 + p3
TOL =~ p4 + p5 + p6
MOOD =~ p7 + p8 + p9
REL =~ p10 + p11 + p12
WITH =~ p13 + p14 + p15
CON =~ p16 + p17 + p18
PROB =~ p19 + p20 + p21

ADDICT =~ SAL + TOL + MOOD + REL + WITH + CON
+ PROB
`

fit <- cfa(VGAmode1, data, estimator="MLM")
```

Example 1: Video game addiction scale

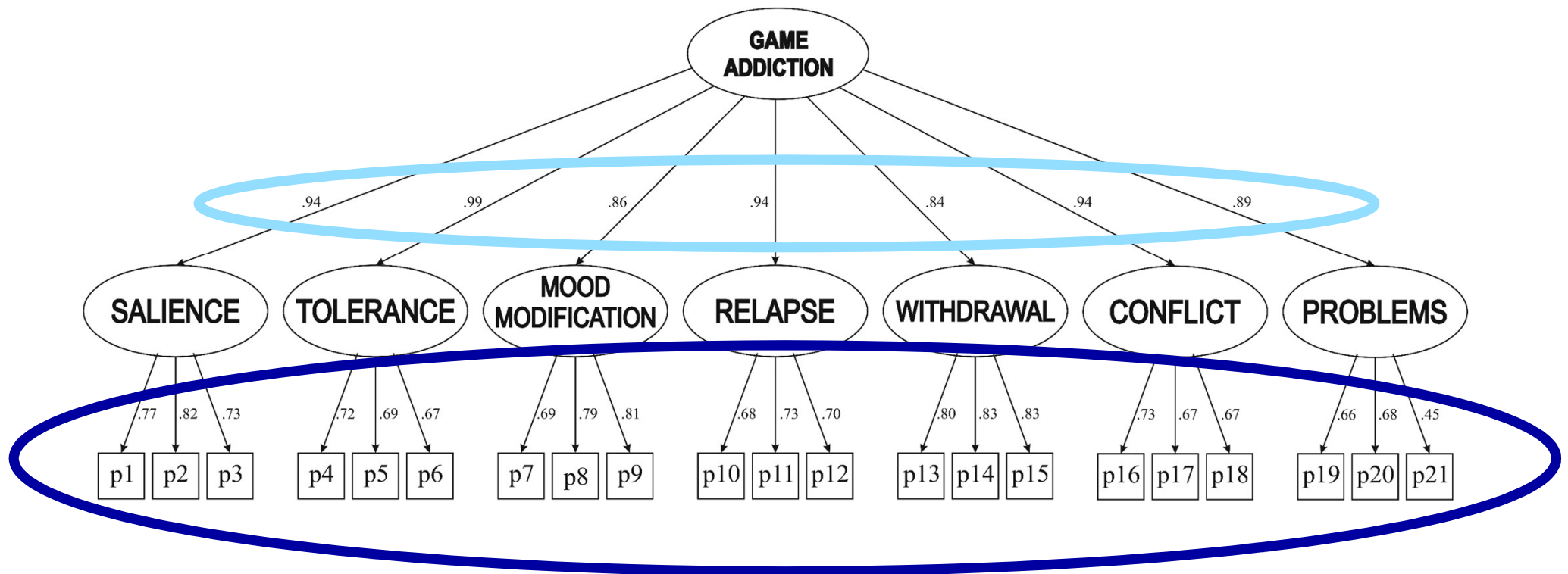
Overall model fit looks good

χ^2/df ratio = 2.4

CFI = 0.90

RMSEA = 0.076

Example 1: Video game addiction scale



Example 1: Video game addiction scale

- Reliability (omega coefficient) of the whole scale > 0.90
- Reliabilities in case of all dimension higher than 0.70 (with the exception of Problems (0.60))

Example 1: Video game addiction scale

As hypothesized, addiction highly correlates with

- Frequency of gaming ($r = 0.65$)
- Game engagement ($r = 0.79$)
- Identity as a player ($r = 0.67$)

Weak relationships were found to

- Quality of sleep ($r = -0.18$)
- Life satisfaction ($r = -0.13$)

Example 2: Children's Depression Inventory (CDI)

Instrument

- Long history of usage, but no generally accepted internal structure
 - Different according to cultural context, unknown in the Czech Republic
- 27 items
- Item format
 - 3 verbally anchored levels of individual symptoms (scored 0 to 2)
 - *I like myself.*
 - *I do not like myself.*
 - *I hate myself.*

Example 2: Children's Depression Inventory

- $N = 1515$ adolescents
- Sample was randomly divided into two parts
 - $N1 = 515$ (used for EFA)
 - $N2 = 1000$ (used for CFA)

Example 2: Children's Depression Inventory

- How many factors in EFA?
- Five indices were computed
 - MAP: 2
 - PA: 4
 - OC: 4
 - AF: 1 (suffers from substantial underextraction)
 - CD: 2
- Correlated or uncorrelated factors?
 - Definitely correlated – use oblique rotation (e.g. direct oblimin)

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, 18(8), 1–14.

Example 2: Children's Depression Inventory

EFA results

Interpretation of factors:

- Based on pattern matrix (with partial correlations between items and factors)
- Core depression symptoms
- Self-esteem
- Performance factor
- Social anhedonia

Example 2: Children's Depression Inventory

```
cdi_4F <- '  
EMO =~ cdi10R + CDI1 + CDI20 + CDI19 + cdi5R + CDI9 +  
CDI6 + cdi8R + cdi16R + cdi2R + CDI3 + CDI27 + cdi18R  
+ CDI17  
EST =~ CDI14 + cdi7R + cdi25R + cdi11R  
EFF =~ cdi15R + cdi13R + CDI23 + CDI26 + cdi24R  
SOC =~ CDI22 + cdi21R + CDI12 + CDI4  
  
DEP =~ EMO + EST + EFF + SOC  
'  
  
fit_all <- cfa(cdi_4F, data, ordered = data[allitems],  
estimator="WLSMV")
```

Example 2:

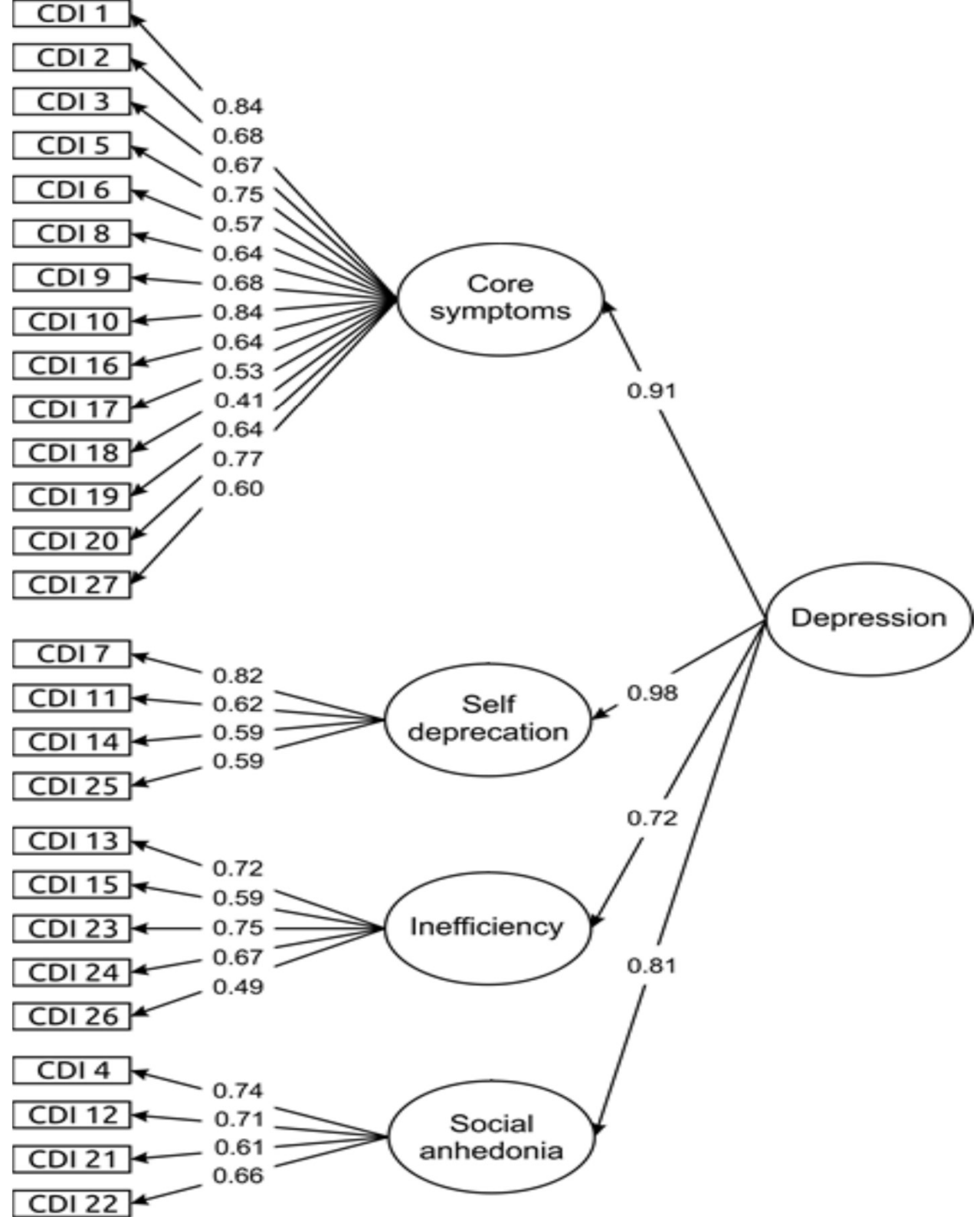
Children's Depression Inventory

Overall model fit

χ^2/df ratio = 4.0

CFI = 0.93

RMSEA = 0.054



References

- Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, 18(8), 1–14.
- International Test Commission (2017). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*. URL: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.

M U N I

Practical exercise

Items construction

Development of a questionnaire: STEP 1:

Identification of constructs that you wish to measure:

Teacher power

- *Power* in the social science context can be understood as an ability of a person or a group to influence opinions, values, and behaviour of others (McCroskey et al., 2006).
- Power is viewed as
 - a *situational* (Jacobs, 2012; Schulz & Oyler, 2006),
 - *circular* (Buzzelli & Johnston, 2001; Aultman, Williams-Johnson, & Schutz, 2009)
 - and *reciprocal* phenomenon (McCroskey, 2006; Moscovici, 2007).

Identification of constructs: 5 teacher power bases

- Traditional and the most influential typology of social power as a *relational* phenomenon comes from French and Raven (1959).
- It distinguishes teacher's power according to the principle which it is based on (as perceived by students).
- The typology of power bases has been developed and partly revised over the years but the main five power bases remained stable (Raven, 1992, 1993).

1. Legitimate power

- *Legitimate power* reflects the teacher's authoritative role in relation to the student.
- Social norms assign to persons who hold position of legitimate authority a certain right to advise or influence others.

2. Coercive power

- *Coercive power* presents a student's awareness that the teacher can punish him/her for example through grade penalties, critique, disciplining in front of classmates, or losing the teacher's favour.
- The teacher power in this case emanates from the student wishing to avoid unpleasantness.

3. Reward power

- *Reward power* comes from a student's perception that the teacher can provide him/her with positive benefits or rewards
 - (extra points, grades, psychological reward such as affirmation from the teacher, relational rewards such as being complimented by the teacher in front of the classmates).
- The teacher power emanates in this case from the student wishing to receive the benefits.

4. Expert power

- *Expert power* emanates from the teacher's knowledge or expertise as an educator in the subject area. In the class, the student may recognize the professional background, superior understanding of the subject, as well as the teaching skills of the teacher.

5. Referent power

- *Referent power* reflects a student's positive regard for the teacher and personal identification with the teacher perceived as similarity or interpersonal affinity being manifested by the student's feeling of unity with the teacher, or the desire to have same identity (i.e. admiring the teacher).
- The teacher's ability to influence a student stems from the positive regard in which the student holds the teacher.

Identification of demographic questions you need to measure

- Characteristics of the sample or population
- Important for your research question
 - You should be able to justify the request (ethical)

Examples:

- Gender/sex: *female male*
- *Total years of teaching:*
- *Taught subject: Mathematics, Science, Czech Language, English, ...*
- *Approbation as a teacher: yes/no*

Development of a questionnaire: STEP 2:

Writing items measuring these constructs

Task 1: Suggest items measuring these 5 power basis

1. Items measuring legitimate power of *student teachers*

- L05: This teacher says that teachers have to be obeyed.
- L07: This teacher thinks that she can decide about everything when she is a teacher.
- L09: When this teacher does not like my behaviour, she cannot do anything about it anyway because she does not belong to our school.
- L11: This teacher emphasizes that we have to obey at school.
- (L14: This teacher has a reserved approach to me.)
- (L17: I obey this teacher because our teacher has told me to do so.)
- (L22: This teacher says that it does not matter if I do not like something in the class.)
- L37: This teacher obviously shows that a teacher is something more than a student.
- (L39: This teacher suggests that what she wants is also supported by our teacher, headmaster or school rules.)
- L42: This teacher says things like: “I end the lesson, not you.”
- L44: When this teacher does not like my behaviour, she cannot do anything about it because she is not a proper teacher yet.
- (L50: This teacher thinks that students have to obey because a teacher is an authority.)

2. Corecive power

- C06: When I do not hand in my homework to this teacher, I feel really bad.
- C16: Although I criticize the rules, this teacher does whatever she wants anyway.
- C18: When I do not work in the class as well as this teacher imagines, she embarrasses me in the class.
- (C25: When I misbehave in the class of this teacher, she tells it to our teacher.)
- C26: This teacher is angry with me when I express myself in the class that I do not agree with what she is saying.
- (C29: When I do not follow this teacher's instructions, she punishes me.)
- C33: When I hand in my homework late, she behaves in such a way it makes me feel bad.
- C34: When I do not work as this teacher wants, she tells our teacher about it.
- C35: When I do not do in the class what this teacher wants, she looks at me angrily.
- C46: This teacher ignores me as a punishment when I do not work as she wants.
- C47: When I do not have my materials for the class, this teacher is upset.

3. Reward power

- RW20: When I know something extra in the class, this teacher points it out.
- RW24: When I work well in the class, this teacher appreciates it.
- RW38: When I behave in the class as this teacher wants, she rewards me.
- RW40: When I work well in the class of this teacher, she tells our teacher about it.
- RW45: When I learn what is required, this teacher praises me.
- RW48: When I make an effort in the class, this teacher is nicer to me.
- RW49: When I do in the class what this teacher demands, she praises me for that.
- RW51: When I behave well in this teacher's class, she praises me to our teacher.

4. Expert power of student teacher

- E02: When this teacher explains something while teaching, it is comprehensible.
- E03: This teacher tells different news connected to the subject.
- E21: I think this teacher is great at teaching.
- E27: When this teacher teaches, I know what to do and when to do it.
- E28: This teacher is able to show me how I can practically use what I learn.
- E30: This teacher understands what she teaches very well.
- E31: When this teacher explains something, I can believe it.
- E36: This teacher is a real expert in this subject.
- E43: This teacher is able to explain to me anything I do not understand.

5. Items measuring referent power of student teachers

- R01: I have a lot in common with this teacher.
- R04: I find this teacher nice because she has to learn as I do.
- R08: This teacher is friendly to me.
- R10: This teacher is fair to me.
- R12: I like to talk with this teacher also during breaks.
- R13: I see this teacher also as a human, not just as a teacher.
- R15: I think of this teacher as of a friend.
- R19: This teacher and I have the same point of view.
- R23: I can see things from the same point of view as this teacher.
- R32: I want to be like this teacher.
- R41: What this teacher says and does is very important to me.

Task 1: Discussion about the suggested items

- Summarize guidelines for the writing of questionnaire items

Guidelines for the writing of questionnaire items

- If items are poorly written, they can create a negative impression on research participants
- => As a consequence, they might complete the items haphazardly or not respond at all

Guidelines for the writing of questionnaire items

- Simply worded items
- Directly relevant items to the construct they were designed to measure
- Begin with a few interesting or nonthreatening items
- Put threatening or difficult items near the end of the questionnaire
- Include examples of how to respond to items that might be confusing or difficult to understand

Guidelines for the writing of questionnaire items

- Avoid terms like *several, most, usually*
 - They have no precise meaning
- Avoid negatively stated items
 - They are likely to be misread by respondents
- Avoid „double-barreled“ items
 - They require to respond to two separate ideas, but only allows a single answer.
 - Example: „*Do you favor abstinence education and drug education?*“
- Avoid biased questions
 - „*What do you see as the benefits of requiring all students to learn algebra in the eighth grade?*“
It assumes that the respondent see benefits when, in fact, they may not.

Evidence of validity and reliability

1. Expert view on items from the point of theory, cultural etc. adaptation of theory
2. Improvements
3. Cognitive interview – how respondents understand the items
4. Improvements of items
5. Pilot study – administration, theoretical structure (EFA/CFA, Cronbach alpha, etc.)
6. Improvements of items, administration etc.
7. Main data collection

References

- Hambleton, R. K., Merenda, P. F., Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah: Lawrence Erlbaum.
- Gall, Gall, Borg (2014). *Applying Educational Research*. Pearson.

M U N I

Any other questions?