Comparability of measurement instruments across countries or time points

Eldad Davidov, Institute of Sociology, University of Zurich

Visit at the University of Tartu, 4.10.12
Content

• 1) Introduction
• 2) The issue of invariance/equivalence in comparative studies
• 3) Data collection and equivalence
• 4) Testing for invariance/equivalence
• 5) Examples from empirical cross-cultural value research
• * 6) Further examples (if time allows) / discussion
1) Introduction: Why comparative studies?

- They help to reveal not only interesting differences between countries and cultures, but also aspects of one’s own country, that would be difficult to detect from domestic data alone.
- They also help to study differences between units other than countries (ethnic groups, ideological groups, professional groups etc.)
- They are also a tool to study change over time.
- Durkheim argued: “Comparative sociology is not a particular branch of sociology: it is sociology itself”. The same is true for other social science disciplines.
• So we need good datasets like the European Social Survey (ESS), which allow us to do this type of work.

• But there is a problem: Human beings have their own value systems, and are capable of saying one thing and doing another thing… They are socially desireable, they are afraid of saying the truth etc.

• So the social sciences have to overcome such difficulties.

• Therefore there is a good reason for great rigour in methodology. Otherwise, results may be very questionable.
• Measuring change over time adds a level of complexity to the analysis and interpretation of findings.

• Measuring cross-national and longitudinal differences and similarities is difficult because of simultaneous variations in many contextual aspects, like social structure, legal systems, politics, economics and culture. These difficulties will not be as severe in a single-country study.

• So cross national studies of attitude change simultaneously incorporate these different difficulties.
• One may even treat cross-cultural studies as **quasi-experiments**. In true experiments there are two groups: an experimental group, which is exposed to a treatment and a control group. Here the **treatment** may be seen as the different culture.

• Whereas in *true experiments* the **treatment** is fully manipulated by the researcher, here it is not. Culture is beyond the control of the researcher. **And-subjects cannot be assigned randomly to culture.**
The ESS as an example

- The ESS provides cross national and over time data about:
- people‘s **value orientations** (values and attitudes to socio-political matters)
- people‘s **cultural/national orientations** (sense of attachment to various groups and their feelings toward outgroups)
- the underlying **social structure** of society (socio-economic and socio-demographic characteristics).
- Can the **questions 'travel successfully'**?
2) The issue of invariance/equivalence in comparative studies

• In other words, we ask **whether our concepts are comparable** across these groups.

• The question we have to ask: does the measurement of our concept change over time? Is it **dependent on the context**: geography, language, legal system, economics, social structure, culture, time point?

• This is especially bothering when we have **subjective variables** like values and attitudes.

• But it is sometimes also troublesome for **objective variables** like education or income.
• To put it simple: when we speak of antisemitism, negative attitudes toward foreigners, authoritarianism, Anomia, values: **is it the same?**
Or: Reflective indicators to measure nationalism and CP: Are they comparable? (ISSP 1995, 2003)
• Or: Being **traditional may mean something else in different cultures** - catholic in Bavaria, religious in Israel etc.

• Items in one culture may be inadequate in another culture. The item 'Do you often **visit your children**' may be adequate in a small country but inadequate/mean a different thing in Germany (as a large country where such often visits are not always possible).
• The term 'immigrants' has a **totally different meaning** in European countries and **in Israel**.

• Europe: **Less rights**, no automatic acceptance into the country, years of **adaptation**, association with social problems, crime, a different group

• Israel: Immigrants are **by definition Jews**. They receive **citizenship upon arrival** with **full rights** and obligations, an absorption basket.
• Hochman, Raijman and Schmidt (2008) find that the item “having a citizenship” to measure attitudes to belongingness is adequate in Germany and not in Israel, and therefore had to be dropped for the cross-country comparison.
• ’Allowing immigrants of a different ethnic group‘ is a problematic question in Hungary.

• It is not as ’neutral‘ as in other European countries, because people associate it with the specific group of Roma people.
Some examples borrowed from J. Harkness

- **US:** Do you have difficulties playing golf? For the Swedish context was changed to: Do you have difficulties picking berries?
- **US:** Running 100 yards Vietnam: Running 3 fields or 100 yards
- **US:** Do you have good appetite? Ethiopia:…?
- **US:** Difficulties reading a newspaper with glasses Africa: who has glasses…

- Historical example: In 1948 international poll, Bulgaria scored highest in ‘do you own a washing machine‘. There was a hand machine that people understood as a washing machine. Translation problem…
- More information in http://www.csdiiworkshop.org
Over time

• The concept of ‘guest workers‘ has completely changed in Germany.

• It used to be a more or less neutral or socially more acceptable a term.

• Today it is absolutely politically incorrect to use it. ‘Guest workers‘ are immigrants who are in Germany to stay…
• Even if an item is **similarly understood**, the scale which is used to measure the item **may be used differently**.

• For example: Why does data on **Greece** in the ESS tend to extremities? How come Greeks are the most traditional, the most loyal to friends, extreme in hostility to immigrants etc…?

• It was found for example that **Hispanics** in the US tend to choose extremes on a 5-point scale, the same with **Mexicans**…

• A ‚yes-saying‘ factor was identified with **Belgian** data among respondents.

• In **Israel** and **Arab countries** one **reads from right to left**. In other countries from left to right. This has to be considered in the **direction of response**.
• Unfortunately, the practice is common to **assume** that instruments (and how they are measured) are equivalent across cultures without any statistical claim to support it.
3) Data collection and equivalence

• **Data collection**: make sure that the data are collected in a comparable manner.

• This is still no guarantee for comparability, so we need some statistical testing whether concepts are comparable.
Comparative Survey Programs: Ideal and Reality (borrowed from W. Jagodzinski)

The Standards

A good Survey should …..

• use accurately translated questionnaires
• consist only of random samples
• display high response rates
• apply always the same mode of interviewing
• be carefully documented
• Include highly reliable and valid measurement instruments

The Reality in many Survey Programs

Actually …..

• Some questions are badly translated
• All kinds of sampling are applied, frequently random route and sometimes even quota samples
• response rates in some countries are low
• Interview modes vary from face-to-face over CAPI, CATI to drop-off questionnaires
• the documentation leaves much to be desired
• The quality of the measurement instruments is dubious and often cannot be examined
Comparative Survey Programs: The Grey Reality

The Survey Program
Actually …..
• Some questions are badly translated
• All kinds of sampling are applied, frequently random route and sometimes even quota samples
• response rates in some countries are low
• Interview modes vary from face-to-face, over CAPI, CATI to drop-off questionnaires
• the documentation leaves much to be desired
• The quality of the measurement instruments is dubious and often cannot be examined

You can ….
… put the survey into trash and take a religious view on the trash problem!
Comparative Survey Programs: The Grey Reality

Religious views on the garbage problem:

• Taoism: Shit happens.
• Hinduism: This shit has happened before.
• Catholicism: If shit happens, you deserve it.
• Islam: If shit happens, it is the will of Allah.
• Judaism: Why does this shit always happen to us?
• Atheism: I don’t believe this shit
• Buddhism: If shit happens, it isn’t real shit.
• Jehova’s Witness: Let me in your house and I’ll tell you why shit happens.
A more pragmatic view: Comparative survey programs as a process

International survey programs have made considerable progress during the last decades. The ESS in particular has set new and higher standards in terms of sampling, data collection, and documentation. However we are still far from perfect!
A more pragmatic view: Dealing with problems of comparability

- We keep in mind that
- **Incompatibilities in sampling and design** may cause **item bias** and the absence of invariance of the scale.
- But **even if** data collection were **comparable**, scales may still be noninvariant.
- Item bias or non-invariance could take place **across countries or over time**.
- Therefore, we always **have to test for it**.
Subjective variables

- Subjective variables are subject to measurement error (probably) more than objective variables.
- They also may be noninvariant
- Therefore we need **reflective multiple indicators** to construct latent variables, where we can control for measurement errors.
- This also enables to **test for invariance/equivalence** across countries and over time of the latent variables/constructs of interests.

- Otherwise we cannot do this. For example: Political orientation is a single question in many survey. It does not allow to control for its measurement errors and test for invariance.
4) Testing for invariance/equivalence

Definition: Measurement invariance refers to “whether or not, under different conditions of observing and studying a phenomenon, measurement operations yield measures of the same attribute” (Horn and McArdle, 1992: 117).
• There are statistical tests to do it.
• The chances to find equivalence of the meaning of questions in the same country across time are much higher than across countries.
• This is because cultures change slowly. People in the same culture use the same language.
• The translation into each nation’s language is a delicate task, where many mistakes can be made.
• Good translation is a necessary but not sufficient condition for invariance. Some idioms may not have an equivalent in another language.
Invariance 2

• Testing for invariance is an important task before starting to analyse cross culturally or over time.

• Generally, there are three steps of invariance: configural, metric and finally scalar invariance.

• Therefore there is a good reason for great rigour in methodology. Otherwise, results may be questionable in cross-country studies (Jowell et al.).
• There is a large set of statistical tools to address construct equivalence. We exemplify the use of multi-group confirmatory factor analysis (MGCFA).

• There is a discussion whether one has to be more strict or more liberal. We come to this later.
Measurement equivalence:

Group A (Culture, country, time point)

\[ \alpha_{1A} \]

\[ \beta_{1A} \]**\[ \beta_{2A} \]

\[ \text{Quest. a} \]

\[ \text{Quest. b} \]

\[ \text{Quest. c} \]

\[ \alpha_{2A} \]

\[ \alpha_{3A} \]

Group B (Culture, country, time point)

\[ \alpha_{1B} \]

\[ \beta_{1B} \]**\[ \beta_{2B} \]

\[ \text{Quest. a} \]

\[ \text{Quest. b} \]

\[ \text{Quest. c} \]

\[ \alpha_{2B} \]

\[ \alpha_{3B} \]
• $X = \tau_X + \Lambda_x \xi + \theta_\delta$

• Where $x$ is the indicator

• $\tau_X$ (tau $X$) is the indicator intercept

• $\Lambda_x$ -Lambda $x$, is the factor loading

• $\xi$-Ksi, the latent exogenous variable

• $\theta_\delta$ -Theta delta, the error variances and covariances

• $M_x = \tau_X + \lambda_x \kappa$

• Where $\lambda_x$ (lambda $x$) is the factor loading, and $\kappa$ (kappa) is the mean of the latent exogenous factor.

• Thus, the mean of a given indicator ($M_x$) can be reproduced by the CFA model's parameter estimates.
• Configural invariance: no constraints
• Metric invariance:
  • where $\Lambda$ (Lambda) stands for the factor loading and $G$ for the group number (country at a specific time point).
• Scalar invariance:
  • Where $\tau$ (tau) stands for the indicator intercept

\[
\Lambda^1 = \Lambda^2 = \ldots = \Lambda^G
\]
\[
\tau^1 = \tau^2 = \ldots = \tau^G
\]
• Configural invariance still does not allow any comparisons.
• Metric invariance allows comparisons of correlates.
• Scalar invariance allows comparisons of means across cultures.
Full vs. Partial Invariance

- Concept of ‘partial invariance’ introduced by Byrne, Shavelson & Muthén (1989)
- It means that only some factor loadings and/or intercepts may be set equal.
- Procedure to test for partial invariance:
  - Constrain complete matrix
  - Use modification indices to find non-invariant parameters and then relax the constraint
  - Compare with the unrestricted model
- Steenkamp & Baumgartner (1998): Two indicators with invariant loadings and intercepts are sufficient for mean comparisons
- One of them can be the marker + one further invariant item
Computation of Means: Equation with Intercept

• $X_1 = \alpha_1 + \beta_1 \times A + e_1$
  ..... (other indicators)

$M(x_1) = \alpha_1 + \beta_1 \times M(A)$

$x_1$ = indicator
$\alpha_1$ = intercept of indicator
$\beta_1$ = factor loading
$A$ = latent variable
$e_1$ = random measurement error of indicator
$M$ = mean
Intercepts and Slopes
equal slopes and unequal intercepts
Exercise

Please draw the following four graphs:
- equal intercepts and equal slopes
- unequal intercepts and unequal slopes
- equal intercepts and unequal slope
- unequal intercepts and equal slopes
• **Invariance** can relate to the **measurements**: factor loadings, item intercepts.

• **Invariance** may also relate to the **structural/theoretical model**: relations between theoretical variables, construct means.
Example

• We measure the value tradition-conformity in the Netherlands, Belgium and Luxembourg with the ESS:
• -It is important to him to be humble and modest. He tries not to draw attention to himself (ipmodst).
• -Tradition is important to him. He tries to follow the customs handed down by his religion or his family (imptrad).
• -He believes that people should do what they're told. He thinks people should follow rules at all times, even when no one is watching (ipfrule).
• -It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong (ipbhprp).
• Questions about measurement:
  – Are the **factor loadings** of the items equal (metric invariance)? It is necessary for comparing correlates (between tradition-conformity and another variable like education) across groups.
  – Are the **item intercepts** in addition equal (scalar invariance)? It is necessary for comparing the mean of tradition-conformity between the Netherlands, Belgium and Luxembourg.
• Questions about structure:
  – Are the relations between tradition-conformity and education equal in the 3 countries?
  – They **may be compared given we find metric invariance**: equal factor loadings of tradition-conformity across the three countries.
  – Without establishing first invariance, such a comparison of the structure is problematic.
TRCO

ipmodst

imptrad

ipfrule

ipbhprp

1

1

1

1

Education
Sequence of model testing in the **full SEM**

- 1) Configural invariance
- 2) Metric invariance
- 3) If metric invariance rejected, partial metric invariance.
- 4) If accepted: add equality constraints on structural effects – to test for differences. This model will include equality constraints on BOTH factor loadings and effects.
• 5) If interested in mean comparisons:
  • Scalar invariance.
  • If the model is rejected, try partial scalar invariance.
  • If accepted, add constraints on means across groups to test for differences. This model will include equality constraints on BOTH factor loadings, intercepts and means.
Second order factor: The syndrome: The Elements in the GFE-Survey (Zick et al. 2008)
2nd order factor: computation

• Same principle as in first order CFA
• Identification problem: Solved by
  – Setting variance of 2nd order factor to 1
    or:
  – Setting one of the factor loadings of 2nd order factor to 1

• How to test for invariance? Only for the first order latent variables? Also for the second-order?
A 2nd order factor model with the values (Schwartz 1992)
How can we deal with item bias?

• Bias can mean that an indicator is inadequate in an instrument for cc comparison.
• As a result: the item may either be removed (often a good solution).
• If an item is removed, the score or meaning of an instrument (construct) may change due to the elimination of the item/s.
• Some of the groups where there is no invariance may be removed from the comparison.
• One could resort to partial invariance if two items are found that are invariant.
• Most radical: the researcher will sometimes refrain from cc comparison.
• Alternatively, item bias can be seen as providing **important clues** about **cross-cultural differences**. One may try to find reasons for the presence or absence of bias.

• It is often difficult to find reasons for item bias…

• This can be done in a **theory-driven** way or **ad hoc** using **multilevel analysis**.
What to do when there is no invariance? (JCCP 2012)

Theoretical expectations:

**H1:** +

**H2:** -
How to decide whether a model is accepted/supported by the data when testing for invariance?

• 1) **History**: Chi square difference test between models that imply different levels of invariance. *Byrne and Stewart (2006)* discuss *problems* with the Chi square difference test. This is called the *traditional/classical approach*.

• 2) **Contemporary** (Chen 2007): *Differences* in CFI between the models. One may also examine differences in RMSEA and SRMR (Chen 2007; Byrne and Stewart 2006). This is called in the literature the *modern/pragmatic approach*.
Model comparison in the pragmatic approach

- As a general rule, Chen 2007 recommends for metric invariance test that if differences in CFI are > .01, for $\Delta$RMSEA > .015, and $\Delta$SRMR are larger than > .03, invariance should be rejected.
- For scalar invariance test:
  - A change of CFI $\geq$ .01, supplemented by a change of RMSEA > .015, or a change of SRMR > .010, would indicate scalar non invariance.
- In general, CFI is good as an over-criterion, since changes in RMSEA and SRMR tend to over-reject an invariant model when sample size is small.
• **The future** (and it is already part of the practice in present studies): Saris, Satorra and van der Veld (2009)

• Only **meaningful** deviations should be considered as violations of invariance.

• The **power** of the test and the size of the **expected parameter change** should be taken into account.

• **JRULE** software was developed to help in the decision making. The **Mplus** version was developed by Daniel Oberski and the **Lisrel** version was developed by William van der Veld.
Willlem Saris       Daniel Oberski

William van der Veld
5) Examples from empirical cross-cultural value research

• Cross-Cultural analysis of the Portrait Value Questionnaire (PVQ) over time and countries: Metric and scalar invariance:
A. Invariance of values with the ESS data

• The three main questions that were asked when analyzing the PVQ data in the ESS were:

(1) **How many** value constructs from the Schwartz theory can be identified in the 20-25 countries included in the 1\textsuperscript{st}, 2\textsuperscript{nd} and 3\textsuperscript{rd} rounds of the ESS?

(2) To what extent are the values **comparable across countries** in spite of diverse cultures and languages?

(3) To what extent are the values **comparable over time** in each country?

• The ESS includes a **shortened version** of the PVQ with **21 double-barelled questions** to measure 10 values.
Structural relations among the 10 value types and the two dimensions
Some examples for the 21 ESS Items

**Power (PO):**
- Important to be rich, have money and expensive things. (Imprich/po1)

**Achievement (AC):**
- Important to be successful and that people recognize achievements. (Ipsuces/ac2)

**Hedonism (HE):**
- Important to have a good time. (Ipgdtim/he1)

**Universalism (UN):**
- Important that people are treated equally and have equal opportunities. (Ipeqopt/un1)
The response categories

Now I will briefly describe some people. Please listen to each description and tell me how much each person is or is not like you.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very much like me</td>
</tr>
<tr>
<td>2</td>
<td>Like me</td>
</tr>
<tr>
<td>3</td>
<td>Somewhat like me</td>
</tr>
<tr>
<td>4</td>
<td>A little like me</td>
</tr>
<tr>
<td>5</td>
<td>Not like me</td>
</tr>
<tr>
<td>6</td>
<td>Not like me at all</td>
</tr>
<tr>
<td>7</td>
<td>Refusal</td>
</tr>
<tr>
<td>8</td>
<td>Don't know</td>
</tr>
<tr>
<td>9</td>
<td>No answer</td>
</tr>
</tbody>
</table>
Findings with the 1st ESS round (Davidov, Schmidt, Schwartz 2008)

• 20 countries, total sample of 39,596.

• 1) How many values were identified?

• 7 values were identified. 3 pairs of values, which are theoretically adjacent, had to be unified, as they were not identified as distinct constructs: Universalism with benevolence, conformity with tradition and power with achievement.
Findings with the 1st ESS round

- Is it a problem?
- This is **not a problem with the theory**, but it is probably an artefact of the type of measurements used (coverage rather than discrimination).

- **Knoppen and Saris (2009)** argue that **replacing several items** with others that can better discriminate between values **will solve the problem**. Their findings were **replicated successfully** by **Beierlein, Davidov, Schmidt, Schwartz and Rammstedt (2012)** using representative German data.
These studies provide important hints how to further develop the measurements:

- They indicate which items have cross-loadings
- which ones have low loadings on their value
- They recommend developing such items that correlate more strongly with items measuring the same value, and correlate more weakly with other items
- now we have new sets of items of Knoppen and Saris (2009), PVQ5X, and PVQ-R.
Findings with the 1st ESS round

• 2) The values identified displayed **metric invariance** across 20 countries.

• So comparisons of relationships seem to be meaningful with these data (Davidov, Schmidt and Schwartz 2008).

• 3) No scalar invariance was supported for the full set of countries, but for small subsets.
Findings with the 2nd ESS round (Davidov 2008)

- **25 countries**, total sample of 47,537.
- In **most countries** (14) the **same 7 values** could be **identified** as in the first round. In the other countries less values could be identified (more values had to be unified).

- **Metric invariance** was **supported** across these countries.
- But – **no scalar** invariance.
Findings with the 2nd ESS round

• What about longitudinal invariance?
• It could be shown that values are (partial) scalar invariant over time in each country.
• In other words, means of values over time may be compared meaningfully, but not across countries.
Findings with the 3rd ESS round (Davidov 2010)

- 19 countries, N=34,904.
- The **findings were similar** to those in the first and second round:
  - 7 or less values could be **identified**.
  - Metric invariance was supported across countries, but **no scalar invariance**.
  - Scalar invariance was found **longitudinally** in each country, thus, allowing the **study of value change** over time in each country.
• **Summary ESS:**
• The scale seems to be too short (too few items) and too much emphasizing coverage to allow identifying more values and finding higher levels of invariance.

• It seems we need a scale with (a) a larger number of items; (b) items that are more similar when measuring a certain value and more heterogeneous when measuring different values (Knoppen ans Saris 2009; Beierlein et al., 2012).

• The **PVQ40** fulfills at least the first requirement.
B. Invariance with the PVQ40 in Germany and Poland (Cieciuch and Davidov 2012)

- 1,204 individuals in Germany and Poland.
- A longer scale with 40 questions.
- All the values displayed configural, metric, and partial scalar invariance with the exception of the stimulation value.
- Thus, the means of nine out of the 10 values may be compared across the samples.
- So we conclude that the full version of PVQ may have better properties for cross-cultural research.
- We argue this carefully, because only two countries were compared.
• This result is not surprising; after all, we have to remember that guaranteeing partial scalar invariance requires having two items with equal loadings and intercepts per construct. With more items per value construct in the full PVQ version, there is a better chance to reach this goal.
Shalom H. Schwartz, The Hebrew University of Jerusalem, Israel and National Research University-Higher School of Economics, Moscow
Jan Cieciuch, University of Finance and Management, Warsaw, Poland
Michele Vecchione, “Sapienza” University of Rome, Italy
Eldad Davidov, Institute of Sociology, University of Zurich, Switzerland
Ronald Fischer, Victoria University, Wellington, New Zealand
Constanze Beierlein, Leibniz-Institute for the Social Sciences, GESIS, Mannheim, Germany
Alice Ramos, Instituto de Ciências Sociais-Universidade de Lisboa, Lisbon, Portugal
Markku Verkasalo and Jan-Erik Lönnqvist, University of Helsinki, Finland
Kursad Demirutku and Ozlem Dirilen-Gumus Atilim University, Turkey
Mark Konty, Berea College, United States
Circle Organized by Motivational Congruence and Opposition

Over 6000 citations of 3 publications that introduced
• The original theory postulated that all values form a **circular motivational continuum** like the color circle
• Values are continuous, no distinct boundaries
• **arbitrarily split into ten** scientifically convenient basic values
• The value circle is organized by **motivational congruence & opposition**
• The broader principles of organization:
  – **personal/social focus** of outcomes (green outer circle)
  – there are **four higher order motivations** (red 2nd circle)
  – values are also organized based on whether they focus on self-protection or growth, on avoiding anxiety or freedom from anxiety
• Subsequent research ignored key assumption that continuum can be split in more fine-tuned values
• More fine-tuned values might yield more precise prediction and explanation
• But: CFA methodological studies of various instruments suggest that even the ten values are often not fully discriminable
• The challenge: Refine the theory, develop instrument that measures more narrowly defined values and that meets test of discriminability
Circle Organized by Motivational Congruence and Opposition
• Theory refinement drew on:
• reexamination of the definitions of the ten basic values (e.g., security—personal & societal)
• examination of all of the MDS analyses of both the SVS and PVQ (300+ analyses)
• looking for spatial discrimination among subtypes of values
• looking at the subtypes suggested by CFA analyses
• Together, it led to split the values into 19 potential values (e.g., 2 PO, 3 UN, 2 CO)
• The idea was to **retain the same motivationally based circle** while
• ordering the 19 values around the circle
• **Self-Direction**: Freedom of thought and action
  – Autonomy of Thought: Freedom to cultivate one’s own ideas
  – Autonomy of Action: Freedom to determine one’s own actions
• **Stimulation**: Excitement, novelty, and change
• **Hedonism**: Pleasure or sensuous gratification
• **Achievement**: Success according to social standards
• **Power**: Control over resources and people
  – Dominance: Authority over people
  – Resources: Wealth and material resources
• **Face:** Maintaining public image [fills gap between SE & PO]

• **Security:** Safety, stability and order
  – Societal: Security in the wider society
  – Personal: Security in one’s immediate environment

• **Tradition:** Maintaining and preserving cultural, family and/or religious traditions

• **Conformity:** Avoidance of violating informal or formal social expectations
  – Rules: Compliance with rules, laws and formal obligations
  – Interpersonal: Avoidance of upsetting or harming others
• Humility: Recognizing one’s insignificance in the larger scheme of things

• Benevolence: Promoting the welfare of one’s in-groups
  – Dependability: Trustworthy and reliable
  – Caring: Devotion to the needs of the in-group

• UN: Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature
  – Concern: Equality, justice and protection for the weak in society
  – Nature: Preservation of the natural environment
  – Tolerance: Acceptance and understanding of those who differ from oneself
### Measuring Values

#### 57 Item Portrait Value Quest. Examples

<table>
<thead>
<tr>
<th></th>
<th>How much like you is this person?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not like me at all</td>
</tr>
<tr>
<td>SD Action: Freedom to choose what he does is important to him</td>
<td>1</td>
</tr>
<tr>
<td>SD Thought: It is important to her to form her own opinions and have original ideas.</td>
<td>1</td>
</tr>
<tr>
<td>BE dependability: It is very important to him to be a dependable and trustworthy friend.</td>
<td>1</td>
</tr>
<tr>
<td>FACE: It is important to her that no one should ever shame her.</td>
<td>1</td>
</tr>
<tr>
<td>HUMILITY: It is important to him to be humble.</td>
<td>1</td>
</tr>
<tr>
<td>CO Interpersonal: It is important to him to avoid upsetting other people.</td>
<td>1</td>
</tr>
<tr>
<td>CO Rules: Obeying all the laws is important to her.</td>
<td>1</td>
</tr>
</tbody>
</table>
# Further Exemplary Items 2

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE—dependability</td>
<td>It is important to him to be a dependable and trustworthy friend.</td>
</tr>
<tr>
<td>BE—caring</td>
<td>It's very important to him to help the people dear to him.</td>
</tr>
<tr>
<td>Humility</td>
<td>It is important to him to be humble.</td>
</tr>
<tr>
<td>CO—Interpersonal</td>
<td>It is important to him to avoid upsetting other people.</td>
</tr>
<tr>
<td>CO—rules</td>
<td>Obeying all the laws is important to him.</td>
</tr>
<tr>
<td>Tradition</td>
<td>He strongly values the traditional practices of his culture.</td>
</tr>
<tr>
<td>SE—societal</td>
<td>Having order and stability in society is important to him.</td>
</tr>
<tr>
<td>SE—personal</td>
<td>He avoids anything that might endanger his safety.</td>
</tr>
</tbody>
</table>
Further Exemplary Items 3

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>Protecting his public image is important to him.</td>
</tr>
<tr>
<td>PO—resources</td>
<td>Being wealthy is important to him.</td>
</tr>
<tr>
<td>PO—dominance</td>
<td>He wants people to do what he says.</td>
</tr>
<tr>
<td>Achievement</td>
<td>Being very successful is important to him.</td>
</tr>
<tr>
<td>Hedonism</td>
<td>Enjoying life’s pleasures is important to him.</td>
</tr>
<tr>
<td>Stimulation</td>
<td>Excitement in life is important to him.</td>
</tr>
</tbody>
</table>
• In a separate study we try to address the questions:
• Can all 19 be distinguished?
• Are they collapsible to 10?
• Are the distinctions useful?
Now we address the issue of cross-cultural Invariance with the PVQ5X across 10 countries (Cieciuch, Davidov, Vecchione & Schwartz, in progress)

• The new scale includes more question items – 57, but also more values to measure – 19. In other words, 3 items are intended to measure each value (Schwartz et al. 2012).
• **Data base** was sample of students and adults in **10 countries**: Finland, Germany, Israel, Italy, New Zealand, Poland, Portugal, Switzerland, Turkey, USA

• Sample size ranged between 141 and 1,295 respondents in each sample.

• Data was collected using either **paper and pencil** or **online** questionnaires. In **some countries** an **11-point scale** was collected in addition to the typical **6-point scale**.

• The report below addresses **findings of the 6-point scale**:
<table>
<thead>
<tr>
<th>Country</th>
<th>Sample</th>
<th>N</th>
<th>Method</th>
<th>Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>Adult</td>
<td>334</td>
<td>P&amp;P 6pt.</td>
<td>M. Verkasalo &amp; K. Porkka</td>
</tr>
<tr>
<td>Germany</td>
<td>Student</td>
<td>325</td>
<td>P&amp;P 6pt.</td>
<td>C. Beierlein</td>
</tr>
<tr>
<td>Israel</td>
<td>Student</td>
<td>394</td>
<td>Online 6pt</td>
<td>Y. Cohen &amp; S. Schwartz</td>
</tr>
<tr>
<td>Italy</td>
<td>2: Adult &amp;</td>
<td>388</td>
<td>P&amp;P 11pt P&amp;P 6pt</td>
<td>M. Vecchione</td>
</tr>
<tr>
<td></td>
<td>Student mix</td>
<td>382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>2: Student</td>
<td>141</td>
<td>Online 6pt</td>
<td>R. Fischer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>527</td>
<td>Online 11pt</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>2: Adult &amp;</td>
<td>545</td>
<td>P&amp;P 6pt P&amp;P 11pt</td>
<td>J. Cieciuch</td>
</tr>
<tr>
<td></td>
<td>Student mix</td>
<td>1295</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student mix</td>
<td>297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>Student</td>
<td>201</td>
<td>Online 6pt</td>
<td>E. Davidov</td>
</tr>
<tr>
<td>Turkey</td>
<td>2: Student</td>
<td>250</td>
<td>P&amp;P 6pt P&amp;P 11pt</td>
<td>K. Demirutku &amp; O. Gumus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>Student</td>
<td>443</td>
<td>Online 11pt</td>
<td>M. Konty</td>
</tr>
</tbody>
</table>
Findings with the PVQ5X

• 1) Single country analyses: All 19 values were identified in each country. It was not necessary to unify any values in these countries. With one exception: Turkey. Translation problems?

• 2) Since the model configuration in Turkey was so different and necessitated unifying values, we conducted the tests of invariance without Turkey.
• 3) The test was conducted on each higher order dimension separately (Cieciuch and Davidov, 2012).

• 4) 9 items that did not load on their respective values properly were excluded from the analyses. Thus, we remained with 48 items measuring 19 values.

• 5) All values displayed configural and metric invariance.

• 6) (Partial) scalar invariance was achieved for most values with the exception of conservation values.
7) We re-ran the models taking into account the ordinal character of the data (as small sample sizes may bias the conclusions) (see Davidov, Datler, Schmidt and Schwartz 2011):

8) The findings could be established again: Values displayed at least partial invariance with the exception of the conservation values tradition, conformity interpersonal, conformity rules and humility.
When we take ordinality into account, things become even more complicated (Davidov, Datler, Schmidt and Schwartz 2011)

**CFA with continuous indicators**

Cont. observed variables

- Latent factor $\eta_1$
  - $\kappa_1, \Phi_1$
  - $\lambda_1 \rightarrow y_1 \rightarrow \delta_1$
  - $\lambda_2 \rightarrow y_2 \rightarrow \delta_2$
  - $\lambda_3 \rightarrow y_3 \rightarrow \delta_3$

**CFA with ordinal indicators**

Latent response variables

- Latent factor $\eta_1$
  - $\kappa_1, \psi_1$
  - $\lambda_1 \rightarrow y^*_1 \rightarrow \delta_1$
  - $\lambda_2 \rightarrow y^*_2 \rightarrow \delta_2$
  - $\lambda_3 \rightarrow y^*_3 \rightarrow \delta_3$

- Ordinal observed variables

- $v_{11}, v_{12}, \ldots, v_{1(K-1)} \rightarrow y_1$
- $v_{21}, v_{22}, \ldots, v_{2(K-1)} \rightarrow y_2$
- $v_{31}, v_{32}, \ldots, v_{3(K-1)} \rightarrow y_3$

$\eta$ (Eta): Latent factor
$\kappa$ (Kappa): Latent mean
$\Phi$ (Phi): Factor variance
$\lambda$ (Lambda): Factor loading
$\tau$ (Tau): Intercept
$\delta$ (Delta): Error variance
$\psi$ (Psi): Factor variance
$\nu$ (Nu): Threshold

$K$: number of categories of the ordinal item

Parameters and latent variables that are specific to the CFA with ordinal indicators in italics
• 9) In other words, we are doing **better** than with previous scales.

• 10) The findings **encourage further improvement of the new PVQ-5X scale (PVQ-R)** and the collection of **representative data** with the new scale.
Conclusions about the invariance studies of values

• The ESS (PVQ21) scale operated well for longitudinal within-country comparisons. However, across countries, it was hard to achieve scalar invariance.

• It seems that PVQ40 operates better than PVQ21 in cross-cultural comparisons, at least for Poland and Germany. Comparisons among a larger set of countries is necessary, to generalize this finding.
• The new PVQ5X scale operates even better than previous scales for cross-cultural comparisons.

• However, even with this scale some values (conservation) do not display the highest levels of invariance, which are necessary for mean comparisons.

• Further assessments of the scale are necessary, especially using representative samples in more (also non-western) countries, to generalize this finding.
• The **PVQ-R** may **overcome** some of the **invariance problems** of the **PVQ5X**, where some **problematic items** were **dropped**. **Increasing** the number of **items** to **57** may **increase the chance to achieve** with the **PVQ-R** at least **partial scalar invariance** for more values.

• Schwartz is even developing now a 3**rd** version with only 2 questions per value (38 in total) with the items that work best (PVQR3).

• GESIS (Rammstedt, Beierlein and Schwartz are developing a short scale (12 items) to measure directly the higher-order dimensions. Similar questions may be addressed with this scale.
Finally, diagnosis and systematic identification of the sources of item bias are an essential step, before measurement scales can be improved for cross-cultural analysis.

This may be done by cognitive pre-/post-tests, but possibly also by theoretically-driven multilevel analyses.
Further Examples

• Does translation ‘kill’ invariance?
• How to explain noninvariance?
• The case of attitudes toward democracy
Study 1

• How Harmful are Translations? A Test with Schwartz’s instrument to measure Human Values.

• With Alain de Beuckelaer, University of Nijmegen, the Netherlands, and University of Ghent, Belgium (IJPOR 2011)
Research Question

• The research question: Are translations a serious obstacle to achieve cross-cultural equivalence of questions in surveys?

• We must ensure that people from different cultures understand the same thing when being asked the same question even if we use a different language.

• One may expect cultures sharing the same language to display higher levels of equivalence compared with cultures with different languages.
• A ‘culturally appropriate translated survey instrument’ satisfies 4 quality criteria (Tourangeau et al. 2000):

  – conceptually equivalent (meaning);
  – technically equivalent (grammar);
  – linguistically appropriate (readable);
  – and culturally competent (reflecting cultural assumptions).

• This is not always easy…
• More than 50 years ago, researchers in the fields of psychology and linguistics already introduced the idea that there are
  – cultural differences in the way people think (cognition)
  – and that this is related with linguistic differences (Whorf, 1956).

• As people think when they reply to a survey, language may be responsible for bias / data non-equivalence, if the translation does not take into account the different ways in which people think.
• How could different languages destroy equivalence?

• It could happen in 4 different stages:

  – comprehension *(attending to the question and deciding what information to search for)*;
  – retrieval *(bringing information from memory)*;
  – judgment *(evaluating the information, judging)*;
  – and response
• Therefore, survey researchers have invested much energy in increasing the quality of translations -
• like in the European Social Survey (ESS).
• Was it successful? This will be tested on the human values scale as measured in the ESS.
Our Empirical Study (and Research Question)

- In the following study we compare pairs of same- and different-language countries.

- **Same language countries are, for example:** Germany-Austria, France – French speaking part of Switzerland, UK – Ireland, Netherlands – Dutch-speaking part of Belgium etc (16 comparisons).

- **Different language countries are, for example:** Germany-France, UK-Netherlands, Austria – French speaking part of Switzerland etc (30 comparisons).
Data and Measures

• The data (on human values questions) are taken from the second and third rounds of the ESS, collected in 2004-2005 and 2006-2007.

• 10 groups:
  – GERMAN (3 groups) : Austria, Germany, German-speaking part of Switzerland
  – FRENCH (3 groups) : French-speaking part of Belgium, France, French-speaking part of Switzerland
  – ENGLISH (2) : Great Britain, Ireland
  – DUTCH (2) : Dutch-speaking part of Belgium, The Netherlands

• Sample size 2004-5 (R2) = 16,915 interviewees;
• Sample size 2006-7 (R3) = 16,992 interviewees
Let us look at three questions to measure values in English, German, French and Italian

1) Universalism:
   - **English**: He thinks it is important that every person in the world should be treated equally.
   - **German**: Er hält es für wichtig, dass alle Menschen auf der Welt gleich behandelt werden sollten.
   - **French**: Il pense que c’est important que tout individu sur terre soit traité de manièrè égale.
   - **Italian**: Secondo lui è importante che tutte le persone in questo mondo vengano trattate allo stesso modo.
2) Tradition:
**English:** Tradition is important to him
**German:** Tradition ist ihm wichtig.
**French:** Les traditions sont importantes pour lui.
**Italian:** La tradizione è per lui molto importante.

3) Hedonism:
**English:** Having a good time is important to him.
**German:** Es ist ihm wichtig, Spaß zu haben.
**French:** Avoir du bon temps est important pour lui.
**Italian:** È importante per lui divertirsi.
Method
Type of analysis:

1) Equivalence test of all possible 16 pairs of 'same language' countries (8 analyses in each round)
2) It was done also for a random sample of 30 pairs of 'different language' countries (15 analyses in each round)
Method of analysis

1) Multigroup confirmatory factor analysis (MGCFA); Differences in global fit measures were used to assess which level of equivalence is achieved (with the program Mplus).

Criticism: the method is too strict

2) Therefore, in addition, a new procedure which was proposed by Saris, Satorra and van der Veld (2009) was applied. It tolerates small/unmeaningful biases and is not so strict. It takes into account the power of the test (with the program Jrule, developed by William van der Veld and Daniel Oberski).
Same language pairs of countries

• In 50% of the pairs (8) at least one value (mostly hedonism) did not achieve partial scalar equivalence.

• In 13% of the pairs (2) 2 values did not reach partial scalar equivalence.

• Dutch- and English-speaking countries always exhibited partial scalar equivalence for all values and rounds.

• French- and German-speaking countries almost never exhibited partial scalar equivalence for all values.
Different language pairs of countries

• In 73% (vs. 50% before) of the pairs (22) at least one value (mostly hedonism) did not achieve partial scalar equivalence.
• In 33% (vs. 13% before) of the pairs (10) at least 2 values did not reach partial scalar equivalence.
• In 20% (vs. 0% before) of the pairs (6) 3 values or more did not reach partial scalar equivalence.
• THIS IS MUCH WORSE THAN BEFORE for same language pairs!
• In other words, a significantly higher percentage of pairs of countries did not reach partial scalar equivalence for all values across different language countries.
Summary and Conclusions

• This study provided some further empirical support for the belief that translations may be (partly) responsible for causing non-equivalence of survey questions across countries.

• So translation could be indeed one source of a problem to achieve equivalence of questions across countries.
• But – there was some indication that it is not only (changes in) language.
  – For English- and Dutch-speaking countries high levels of equivalence were generally established for all values
  – German- and French-speaking countries tend not to demonstrate such high levels of equivalence for all human values.

• It could indicate a larger cultural distance among French- and German-speaking countries/parts of countries when we compare this to Dutch-speaking or English-speaking countries in Europe.

• More units of analysis may be necessary to study this aspect more closely.
• **Limitations**: The quasi-experimental nature of our study did not enable us to control for cultural distance.

• Even though human values are important determinants of attitudes, individual, and organizational behavior, other variables are important as well (e.g. the Big Five).
Study 2

Eldad Davidov, University of Zurich
Hermann Dülmer, ZA of the University of Cologne
Elmar Schlüter, WZB - Social Science Research Centre, Berlin
Peter Schmidt, University of Marburg
Bart Meuleman, University of Leuven
• The purpose of the study: If invariance is absent, then: **Explaining non-invariance of one value – universalism**, by using multilevel analysis in a theory-driven way.
Schwartz’ Items on Universalism in the ESS (male version):

- **Equality**: He thinks that it is important that every person in the world be treated equally. He believes everyone should have equal opportunities in life.
- **Understanding**: It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them.
- **Environment**: He strongly believes that people should care for nature. Looking after the environment is important to him.
Empirical Results of the Invariance Test

- Lower levels of invariance: configural and metric invariance, were confirmed by the data
  - Factor loadings are equal across countries.
- Scalar invariance rejected
  - Indicator intercepts are different.
- More specifically: The intercept of the indicator Environment is different across countries.
- Not only methodological artefact.
Empirical Results of the Invariance Test

What to do?

1) One could follow Byrne, Shavelson and Muthen (1989) and use universalism for cross country studies. The reason: it is partially invariant. As long as at least two indicators are invariant, the concept may be used in cross cultural studies.

2) The second strategy would be to try to explain why invariance is absent for the environment item.
• Inglehart (1990, 1997) argues that one places the greatest subjective value on things that are relatively short in supply (the scarcity hypothesis).

• His second argument suggests that one’s basic values reflect the conditions that prevailed during one’s pre-adult years (the socialization hypothesis).
Based on these arguments of socialization and scarcity, Inglehart explained postmodernization:

Higher-order postmaterialist needs (reflected in universalism) are assumed to take priority over more fundamental materialist needs for economic and physical security when such materialist needs are fulfilled. Such a shift took place after World War II in West European societies.
According to these argument, we expect Schwartz’ value of universalism (which reflects the total of three elements, environmental protection, equality, and tolerance toward different people) to be more important in postmodern advanced industrial countries than in less developed modern countries (H1 – the socialization hypothesis).
Inglehart also states that in less developed countries where air and water pollution are far worse than in advanced industrial societies, environmental protection is overall less a quality of life issue but much more a matter of physical health (scarcity).

In such societies people are more likely to support improved environment protection (cf. Inglehart 1997:242).

Therefore, it can be expected that (improved) environmental protection (as a single component of universalism) is perceived as more important in less developed modern countries than in postmodern advanced industrial countries (H2 – the scarcity hypothesis).
Macro Data

Measurement for a Society’s Economic/Technical Development: HDI

The Human Development Index (HDI) combines

- a country’s standard of living (GDP pc in PPP $US),
- educational attainment (combined index for adult literacy rate and combined gross enrollment rate in primary, secondary and tertiary schools),
- level of longevity (life expectancy at birth)

Observed Range of the HDI 2004:
- Turkey: 0.757
- Norway: 0.965
One-Level CFA versus Two-Level CFA

Country Level: Between

Respondent Level: Within

AIC: 368050.207
BIC: 368171.824
Adj. BIC: 368127.332
RMSEA: .003
SRMR (within): .000
SRMR (between): .062
What to do when there is no invariance?

Theoretical expectations:

H1: +

H2: -
Explaining Non-Invariance in Universalism via Multilevel SEM

Country Level: Between

HDI 2004

-2.965

1.165

Environment (impev): 4.680

Underst. Diff. People (ipudrst): 2.613

Equality (ipeqopt): 2.830

Environment (impev)

.711

.735

.334

.960

1.096

Equality (ineqopt)

Underst. Diff. People (ipudrst)

.735

1.069

Underst. Diff. People (ipudrst)

.024

1.747

Underst. Diff. People (ipudrst)

.015*

1.000

Universalism Between

.000

.703

Universalism Within

AIC: 368042.483
BIC: 368181.474
Adj. BIC: 368130.625
RMSEA: .000
SRMR (within): .000
SRMR (between): .045
Study 3: The case of attitudes toward democracy (with Gal Ariely, Univ. of Haifa, SIR, 2010)

Confirmatory Factor Analysis Model of Democracy-Autocracy Preference (DAP) Scale

Having a strong leader (V 164)  →  e1
Having experts (V165)  →  e2
Having the army rule (V166)  →  e3
Having a democratic political system (V167)  →  e4

Source: Data are taken from the WVS 2000, N=51067.
• Invariances test **collapses** when all the WVS countries are included (scale is used in comparative studies with invariance assumed).

• However, a different scale works much better: evaluation of performance of democracy.
The Democratic Performance Evaluation Scale

**Source:** Data are taken from the WVS 2000, $N=51067$. 
Concluding points
(Van de Vijver and Leung 1997)

• Statement 1) Cross-cultural differences in scores on social and behavioral measures tend to be open to multiple interpretations: Are differences due to ’real‘ differences? OR due to different understanding of the concept? Methodological artefact like different scale use? Translation problem? Inadequate items?

• Statement 2) Cross-cultural studies should assess in each group the appropriateness of: the constructs examined, and their operationalizations.
• Statement 3) **know the cultures, check if items apply** to all of them, etc. (can a similar question to measure attitude toward democracy apply in Iran and in Germany?).

• Statement 4) The decomposition of culture into **context variables** is a methodologically powerful means to **provide interpretations for cc differences**. One could consider multilevel techniques.
• **Statement 5)** Equivalence/invariance should be established and cannot be assumed.

• **Statement 6)** Higher levels of equivalence are more difficult to establish (metric invariance is more difficult to establish than configural invariance; scalar more difficult than metric).

• **Statement 7)** In sum: For optimization of the interpretability of cross-cultural differences and similarities we need a combination of substantive, methodological and statistical considerations.
• Statement 8) CC research is essential in establishing the **generalizability of theories and empirical results**. All the more, when subjective variables are involved.
• Thank you very much for your attention!!!!
Appendix

• In which type of studies is invariance relevant?
4 common types ('ideal types') of cross cultural (cc) studies*

<table>
<thead>
<tr>
<th>Consideration of contextual factors</th>
<th>Orientation more on</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothesis Testing</td>
</tr>
<tr>
<td>No</td>
<td>Generalizability</td>
</tr>
<tr>
<td>Yes</td>
<td>Theory-driven</td>
</tr>
<tr>
<td></td>
<td>Exploration</td>
</tr>
<tr>
<td></td>
<td>Psychological differences</td>
</tr>
<tr>
<td></td>
<td>External validation</td>
</tr>
</tbody>
</table>

* From van de Vijver and Leung, 1997
Two dimensions

• 1) Exploratory vs. Hypothesis-driven studies: In exploratory studies, researchers do not have firm ideas about the cross-cultural similarities and differences to be expected (for example studies about unknown cultures or insufficient previous research to generate hypotheses).
2) The second dimension: addressing the use of context variables to explain cultural similarities or differences. Some studies do not use them and the goal is only to document cultural similarities and differences.

In reality studies are a mix of these categories.
The 4 types of studies:

- **Generalizibility studies** (hyp, no context. var.): establish generalizability of research findings in one group (for example western) to other groups (for example non-western). There is no reference to cultural elements that explain the differences (because one wants to generalize).
• **Theory-driven studies** (hyp., contextual var.): specific aspects of a culture such as ecology, or cultural habits are part of the theory.

• Contextual variables are used to explain differences (or similarities).

• Often multilevel analysis is used.
• **Psychological differences** (no hyp., no contextual var.): very common. A *measurement instrument* is applied in at least 2 cultures (Germany and Austria), and the researcher is interested to **compare** them: means, standard deviations, reliability and structure. **No theory for differences** or expectation about any differences that may be found. **Post-hoc explanations** for differences to interpret the differences.
• **External validation studies** (no hyp., contextual var.): one can do the same thing with context variables:

• It is an attempt to explore the meaning and causes of cross-cultural differences with the aid of context variables.

• There are **no a-priori hypotheses**. A large set of **context variables** is used in an exploratory manner.
Two general orientations in cc research: level and structure

- **Structure-oriented studies**: focus on relationships among variables, and attempt to find similarities and differences in those relationships across cultures.
  - For example: is the effect of values on attitudes equal across European countries? Is the structure of values similar across cultures?

- **Level-oriented studies**: focus on differences in magnitude of variables across cultures. For example: do individuals in Austria show higher levels of tradition values compared with Germans?

- Each of the 4 types of cc studies can adopt either a structure or a level orientation (or both).
Examples

• **Generalizibility studies** (hyp, no context):
• Schwartz’s 1992 work on the universality of the structure of values (structure orientation): The hypothesis was the generality and universality of his values.
• **Theory driven studies** (hyp, context):

• Semyonov, Raijman and Gorodzeisky (2006) expected the **size** of immigrant population and **economic conditions** to affect the level of **negative attitudes** toward immigrants.
• **Psychological differences studies** (no hyp, no context):

• Russel and Sato (1995): studied the meaning and *equivalence of emotion words* among English-speaking and Japanese-speaking individuals.

• **Level-oriented**: Guida and Ludlow (1989) studied the cross-cultural *differences* in level of *anxiety* among American and Chilean school children.

• Or *differences in importance of certain values across countries.*
• **External validation** studies (no hyp, context):

• **Structure-oriented**: Diener and Diener (1995). Relations between life satisfaction and satisfaction with family, friends and finances were studied. CC differences in these correlations were explained by the level of individualism/collectivism of the society (ad hoc).

• Davidov, Meuleman, Billiet and Schmidt (2008) explained ad hoc differences in the effect of universalism on support for immigration using economic conditions and inflow of immigration as context.
• In the initial stage, when not much is known about how a phenomenon varies across cultures and why, psychological differences and external validation studies (no hyp, exploration) will help to identify patterns of cross-cultural similarities and differences.

• When a sufficient number of empirical studies are available, theory-driven and generalizability studies can be designed to test these statements.