

Questionnaire Design for Cross-Cultural Surveys WAPOR Training Workshop

Lydia Repke

WAPOR Conference 2024, Seoul, South Korea 28 July, 2024



Introduction

Getting You Settled In





+ More Guides (Scan the QR code)

How to use:

1 Phone

WiFi – SKKU_SEMINAR – Input Key (skku0728)





 \bullet \bullet \bullet \bullet \bullet

Introduction

Welcome





OR

Please enter the code **2917 2101** and answer the questions.





Introduction

Outline

- Aims and Types
- Challenges of Cross-Cultural Surveys
- Measurement
- Techniques and Tools
 - Cognitive Pretesting
 - Web Probing
 - The Survey Quality Predictor (SQP)
- Translation



4





Why Compare Countries (Cultures)?

- Test the generality of theories (micro level)
 - If $x \rightarrow y$ in the US, $x \rightarrow y$ everywhere? Same strength?
- If not: Investigate contextual effects (macro level)
 - Compositional effects (e.g., demography)
 - Economic effects (e.g., level of development)
 - Socio-structural effects (e.g., inequality)
 - Institutional effects (e.g., political system)
 - Historical reasons





Why Compare Countries (Cultures)?

- Cross-national analysis is an essential part of the development process of social-science theory
- Development and testing of statements concerning relationships between variables
- Differences between countries (cultures) are interesting in themselves



Cross-Cultural Surveys

- Definition: multipopulation surveys that are deliberately designed for comparative research
 - \rightarrow to produce comparable data across populations (i.e., countries, cultures, languages)
- Aim: to quantitatively compare descriptive statistics and multivariate relationships between populations



Types of Cross-Cultural Surveys

- Monolingual surveys may still be cross-cultural when covering multiple ethnic/migrant groups
- Multilingual surveys may include the official languages in a country and/or languages of migrant groups

	1	country
1+ countrie	1+	countrie



	1 language	1+ languages
	Monolingual survey (cross-cultural)	
es	Cross-national surveys (cross-cultural)	

 Cross-national and multilingual surveys are always cross-cultural

Examples of *Cross-National* **Surveys**

- Academic
 - International Social Survey Program (ISSP)
 - European Social Survey (ESS)
 - World Value Survey (WVS), European Values Study (EVS)
- Governmental/European Union
 - Eurobarometer (EB)
 - European Labour Force Survey (EU-LFS)
- International Organizations •
 - Programme for International Student Assessment (PISA)







Aims and Types

Examples of *Multilingual National* **Surveys**

- <u>Swiss Household Panel</u> (Switzerland)
- IAB-SOEP Migration Sample (Germany)



11

But the Methodological Question Is

Are the data truly equivalent across countries, regions, languages, and cultures so that we can analyze and **compare them?**

What are

- real differences,
- real similarities, and
- biases that reduce or completely hinder comparability?



12









Challenges in Cross-Cultural Research

- **Comparability of data**
 - Measurement: Conceptual and indicator issues, multiple languages and difficulties of meaning
 - **Representation:** Obtaining equally good sampling frames everywhere
- Large coordination/organizational efforts required, demanding and expensive
- Standards, legal/ethical concerns, practical problems and expertise in survey design and data collection differ by country



Comparability of Measurement

- Nominally comparable questions at the level of language might be understood differently in different locations
- If questions and indicators are not functionally equivalent for the purpose of analysis, we cannot justifiably compare the data collected
- Goal for analysis: one single data file for all countries with largely identical variables



Types of Biases

- **Construct bias:** the construct across cultures is not identical (van de Vijver, 1998)
- Method bias: biases due to the method or measurement context • Sample bias: cross-cultural variations in sample characteristics Instrument bias: differences in response style, e.g., acquiescent or
 - extreme (Johnson et al., 2011)
 - Administration bias: differences in data collection modes/procedures (e.g., face-to-face vs. web)
- Item bias: poor item translation, ambiguous source items, inapplicability of item contents or connotations associated with the item wording (*He & van de Vijver, 2012*)



Example of Construct Bias

Filial Piety

= psychological characteristic associated with being "a good son or daughter" (Ho, 1996)

Chinese Concept

- Children should assume role of caretaker of elderly parents
- Social aspects (e.g., obedience)
- Broader than the Western concepts





https://de.wikipedia.org/wiki/Kin dliche_Piet%C3%A4t

Example of Item Bias

Religious Involvement (Billiert, 2013)

- ESS round 2, 3 items
 - Regardless of whether you belong to a particular religion, how religious 1) would you say you are?
 - Apart from special occasions such as weddings and funerals, about how 2) often do you attend religious services nowadays?
 - Apart from when you are at religious services, how often, if at all, do you 3) pray?
- Lack of comparability in Turkey (Muslim-majority country)
- Islam: not customary for females to attend religious services



Example of Item Bias

"If immigrants commit serious crime they should be made to leave" (ESS round 1)

- Denmark the most tolerant country in ESS data
- Unlikely: electoral success of the Danish extreme right-wing party (Coenders et al., 2008)
- "Tolerance" of Danish introduced by translation (Billiet, 2013) Danish translation had larger lexical scope and also included minor crimes or offenses (e.g., violation of traffic rules)
- Term "serious crime" usually associated with homicide and rape (Schulz, Meitinger, Braun, & Behr, 2018)



Comparable Measurement in Cross-Cultural Research

- ... requires ex-ante procedures to ensure equivalence
 - adequate cross-cultural source questionnaire development and translation
 - comprehensive cross-cultural testing
- ... requires ex-post procedures to generate and assess equivalence
 - statistical procedures prior to substantive analysis (e.g., Braun & Johnson, 2010; Davidov et al., 2014)
 - output harmonization



equivalence re development and

e and assess equivalence nalysis (e.g., Braun &

Methods for Comparable Measurement

Harmonization

- adjusting data collection with the aim to achieve comparability (wide definition)
- constructing comparable variables out of country-specific variables (narrow Definition)
- Translation: transforming questionnaire items from one language/culture to another
- Adaptation: introducing differences in items to achieve equivalence of meaning









The "Translation" Process

questions?

How can theoretical concepts/constructs "translated" (transformed or transferred) into





be



The "Translation" Process

Three Steps

- Specify the concept and distinguish between concept by intuition and concept by postulation
- 2) Operationalize the concept
- 3) Measure





Two Types of Processes

Concepts by Intuition

are simple concepts whose meaning is immediately obvious. *e.g., evaluations, importance, demographic variables*

Concepts by Postulation (Constructs)

are less obvious concepts that require explicit definitions. They are also referred to as constructs. They cannot be measured by single items but by multiple items representing concepts by intuition.

e.g., racism, religion, integration, power







Blalock, 1990

From Concept to Measurement

Concept of interest

theoretical term, not directly observable

Concept specification

defining the concept, we want to measure and specifying orientation and the involvement
its dimension(s)
orientation and the involvement
with religion

Operationalization

translating the concept into a set of specific operations and procedures that are suitable to empirically capture the concept's meaning



Measurement

assigning numbers to objects according to defined rules in a way that a structurally accurate mapping is created



(e.g., religiosity)

(religiosity = religious

(frequency of church attendance, frequency of praying, type of religion believed in)

> (low/high values reflect low/strong religiosity)

Let's Say You Want to Measure...



alcohol consumption in Europe

Alcohol Consumption in the ESS

Measurement according to the WHO





= consumption in equivalent liters of pure alcohol (ethanol) per capita



Alcohol Consumption in the ESS

"Please think about the last time you were drinking alcohol on a Monday, a Tuesday, a Wednesday <u>or</u> a Thursday. How many of each of the following drinks did you have on that day?

Use this card to guide your answer. Any other drinks?"





Alcohol Consumption in the ESS





 \bullet \bullet \bullet

 $\bullet \bullet \bullet \bullet \bullet \bullet$

 $\bullet \bullet \bullet \bullet \bullet \bullet$

Making Alcohol Consumption Comparable



n = number of alcoholic beverages k = type of alcoholic beverage



32

Binge Drinking in the ESS

Measurement according to the National Institute on Alcohol **Abuse and Alcoholism in the United States**





= a pattern of drinking that brings a person's blood alcohol concentration to 0.08 gr % or above

Binge Drinking in the ESS

This card shows six different examples of how much alcohol a person might drink on a single occasion. In the last 12 months, how often have you drunk this amount of alcohol or more on a single occasion? Was it...

- daily or almost daily,
- weekly,
- monthly,
- less than monthly,
- or never?



Beispiel 1	Beispiel 2
3 große Gläser Bier, 2 kleine Gläser Bier	3 große Gläser Wein, 1 kleines Glas Wein
Beispiel 3	Beispiel 4
4 Cocktails/Longdrinks, 1 Glas Likör,	3 große Gläser Bier,
2 Gläser Spirituosen	3 Gläser Spirituosen
Beispiel 5	Beispiel 6
2 große Gläser Wein, 2 Cocktails/Longdrinks, 1 Glas Spirituosen	2 Flaschen Bier, 2 Cocktails/Longdrinks, 3 Gläser Spirituosen

Binge Drinking in the ESS

Beispiel 1	Beispiel 2
3 große Gläser Bier	2 große Gläser Wein, 1 kleines Glas Wein
Beispiel 3	Beispiel 4
3 Cocktails/Longdrinks, 1 Glas Likör, 1 Glas Spirituosen	2 Flaschen Bier, 2 Cocktails/Longdrinks
Beispiel 5	Beispiel 6
1 Glas Sekt, 2 kleine Gläser Wein, 1 Glas Likör,	2 Flaschen Bier, 1 Cocktail/Longdrink,
1 Glas Spirituosen	2 Gläser Spirituosen







 $\bullet \bullet \bullet \bullet \bullet \bullet$



 $\bullet \bullet \bullet$

Binge Drinking in the ESS





Beispiel 1	Beispiel 2
3 große Gläser Bier, 2 kleine Gläser Bier	3 große Gläser Wein, 1 kleines Glas Wein
Beispiel 3	Beispiel 4
4 Cocktails/Longdrinks.	
1 Glas Likör, 2 Gläser Spirituosen	3 große Gläser Bier, 3 Gläser Spirituosen
Beispiel 5	Beispiel 6
2 große Gläser Wein, 2 Cocktails/Longdrinks, 1 Glas Spirituosen	2 Flaschen Bier, 2 Cocktails/Longdrinks, 3 Gläser Spirituosen

 $\bullet \bullet \bullet$

 \bullet \bullet \bullet \bullet \bullet


Techniques and Tools

Goal of Question Evaluation and Pretesting

Survey pretesting refers to the evaluation of individual questions or the whole questionnaire before their use in the field.

- Evaluating the adequacy of questions and materials
 - for the intended purposes
 - for the intended population
- Identifying and analyzing sources of measurement error in questions
- Offering recommendations for improvements
- Uncovering problems with translated and adapted survey items
- Ensuring validity and cross-cultural equivalence



Methods for Testing Survey Questions

- Standard Pretest (Pilot study)
- **Expert Review**
- **Behavior Coding**
- **Respondent Debriefing**
- Focus Groups
- **Cognitive Interviewing**
- Eye Tracking

- SQP

. . .



Web Probing **Split-Ballot Experiments** MTMM Design



Techniques and Tools

Cognitive Interviewing

"...the administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help **determine** whether the question is generating the information that its author intends."



Beatty & Willis (2007:278)

Key Elements of Cognitive Interviewing

1) Administering the survey questions

Participant answers the question that should be tested

2) Observation and think-aloud

- Interviewer observes reactions during question answering
- Problems are documented if they show up at this point already

3) Probing

One survey question after the other is discussed, specified follow-up questions (probes) are asked







Techniques and Tools

Key Elements of Cognitive Interviewing

1) Administering the survey questions

2) Observation and think-aloud

3) Probing



Qualitative information is collected on

- understanding of words
- recall and memory
- estimation strategies
- comfort level with answering
- confidence in accuracy of answer

Conducting Cognitive Interviews

- One-on-one interview (usually in the participant's home)
- Practice ranges from (completely) unstructured to (completely) standardized interviewing
- Usually small number of interviews (10-30 cases)
- Participants should resemble the respondents of the actual survey with regard to sex, age, education, and other important study-specific variables
- Conduct the pretest in the same mode of data collection as the main survey (interviewer administered or self-administered)



Procedures for Cross-Cultural Cognitive Interviewing

- Ideally, should include source language testing (Carter, Schoua-Glusberg, & Sha, 2009)
- Ideally, pretest instruments in all populations
 - Check adequacy in all cultures
 - Check interpretation of questions in all cultures
 - Check possible sources of error in all cultures
 - Pretest part of translation method TRAPD





Procedures for Cross-Cultural Cognitive Interviewing

- Each additional country/culture/language provides valuable information but is increasing the complexity of the pretest
- Cognitive interviewing across countries is very difficult to organize
 - Availability of cognitive laboratories
 - Standardization of procedures





Techniques and Tools

Phasing of Testing

1) Sequential testing

Source questionnaire first, refinements, and then test in other countries, cultures, languages

2) Parallel testing

Source questionnaire tested simultaneously in a number of different countries, cultures, languages

Ideal: Parallel development of source and different language versions, allows two-way feedback



Cognitive Interviewing

Cognitive pretesting is especially important for cross-cultural surveys.

Issues covered in cross-cultural research

- Lack of comparability; different interpretations
- Differences in level of underlying knowledge
- Differences in understanding: translation, cultural background
- Differences in naturalness of language
- Differences in answer scale use
- Lack of construct overlap between source and target questionnaire



Techniques and Tools

Web Probing

Goals

- Reveal respondents' cognitive processes when answering a survey question
- Assess the quality of survey questions
- Cross-cultural studies: uncover equivalence problems in cross-national surveys and identify different answer patterns across countries

Method: Application of probing techniques from cognitive interviewing in web surveys







When you answered the previous question about your health, what did you think of?





How would you rate your health in general? Would you say

When you answered the previous question about your health, what did you think of?

nditions Closed probe **Techniques and Tools**

Web Probing Design

Adapt size to the expected answer





50

Strengths for Cross-Cultural Testing

Sampling & Recruitment

- Larger sample sizes with a broader geographical coverage
- Time- and cost-efficient recruitment
- Relatively fast data collection

Mode

- No need for (availability and training of) cognitive interviewers
- Higher standardization across subjects, self-administration rules out interviewer effect
- No transcription of interview data needed

Behr et al., 2017; Lenzner & Neuert, 2017





Strengths for Cross-Cultural Testing

Analysis

- Findings can be quantified
- Rules out "false positives"
- Around 100-120 answers sufficient per language/country to obtain meaningful results

Behr et al., 2017; Lenzner & Neuert, 2017





Weaknesses for Cross-Cultural Testing

Mode

- Lack of motivating interviewer
- No spontaneous or conditional probing
- Only scripted probes possible, no follow-up on incomplete answers ("one-shot")

Response Quality

- Higher amount of "probe nonresponse" and mismatching responses
- Shorter answers, not interpretable answers
- Higher response burden

Behr et al., 2017; Lenzner & Neuert, 2017



53

Weaknesses for Cross-Cultural Testing

Analysis

- Higher effort in data analysis due to larger sample size
- Translation of probe responses





Web Probing Example

National Identity (items from the ISSP 2013)

- How proud are you of [COUNTRY] with regard to its social security system?
- Wie stolz sind Sie auf Deutschland hinsichtlich sozialstaatlicher Leistungen?
- ¿Qué tan orgulloso/a está Ud. de México con respecto a su sistema de seguridad social?
- Finding: 39% of web probing respondents were thinking of the security situation in Mexico (referred to as silent misunderstanding)









✓ Resources

WHAT IS SQP?

SQP 3.0 is a survey quality prediction system for continuous questions used in survey research.

LATEST NEWS

The Survey Quality Predictor

Sneak Preview

Check out our preprint.

Read up on the development of SQP 3.0's prediction algorithm.

SQP 3.0 @ WAPOR 2024 (July 28-31)

Training Workshop on "Ouestionnaire Design for Cross Cultural Surveys"



- ✓ CONSULT, COMPARE, AND EVALUATE CONTINUOUS QUESTIONS
- ✓ DESIGN NEW QUESTIONNAIRES
- ✓ CORRECT FOR MEASUREMENT ERRORS

↗ read preprint



But what Exactly Is SQP?

The Survey Quality Predictor (SQP) is

- A web-based software to estimate the measurement quality of survey items based on their characteristics (e.g., scale properties)
- A searchable database of existing survey items with information on their measurement quality
- A coding system to compare different language versions of survey items



Background

Saris & colleagues started conducting and analyzing a series of MTMM experiments

SQP 1.0 for Windows Saris, Oberski, & Kuipers, 2004

1990's

Idea of SQP

Hands-on tool for predicting the quality of *new* questions based on the questions' characteristics

2000's SQP for DOS *87 MTMM* experiments in English, German, and Dutch *Saris, 2001*



AAPOR's Warren J. Mitofsky Innovators Award (2014)

SQP 3.0 ~ 600 MTMM *experiments; 33 countries; 28* languages Felderer et al.

2010's SQP 2.1 (2015)

2020's

SQP 2.0 + >250 MTMM experiments from the ESS; 22 countries; 21 languages *Saris et al., 2011*









Formulation

Response scale

Additional components (introduction, data collection mode, etc.)

Item Characteristics Influence the Response Behavior

Language & Country of Survey

Trait

- Domain
- Concept

Associated to Trait

- Social Desirability
- Centrality
- Reference period

Instructions

- Interviewer instruction
- Respondent instruction

Additional Information

- Knowledge provided
- Extra information or definition

Introduction

Request in intro

Linguistic Complexity

- # sentences
- # subordinated sentences
- # nouns
- # abstract nouns
- # syllables

Formulation of Request

- Direct, indirect, no request
- Stimulus (battery)
- WH word used in the request
- Type of 'WH' word
- Interrogative, imperative, declarative
- Gradation
- Balance
- Encouragement to answer
- Emphasis on subjective opinion •
- Information about the opinion of other people
- Absolute or comparative judgment

Response Scale

- # categories
- Full/partial labels
- Long/short labels
- Order of labels
- Label-number correspondence
- Theoretical range (bipolar/unipolar)
- Range used
- Symmetry of response scale
- Neutral category
- Number of fixed reference points
- Don't know option



- Categories, yes/no scale,
 - frequencies, magnitude
 - estimation, line production...

Visual Aid Characteristics

- Showcard or other visual aids used
- Horizontal or vertical scale
- Overlap of scale labels and categories
- Start of the response sentence
- Request on the visual aid
- Numbers or letters before the answer categories
- Scale with only numbers or numbers in boxes
- Picture provided?

Questionnaire Context

- Computer-assisted
- Interviewer
- Visual or oral presentation
- Position of item

Why Do We Have to Bother?

Because all these decisions have an effect on the response behavior and as such also influence the quality

Measurement quality/measurement error







Estimating the Quality of Survey Items

Multitrait-Multimethod (MTMM) Model

- Involves: measuring several related concepts by asking them multiple times using slightly different methods
- Design: 3-trait-by-3-method model (most common)

Trait 3: democracy

Method 1: 0 (extremely dissatisfied) – 10 (extremely satisfied) Method 2: 1 (very dis.), 2 (fairly dis.), 3 (fairly sat.), 4 (very sat.) Method 3: 0 (extremely dissatisfied) – 5 (extremely satisfied)

Campbell & Fiske (1959); Saris & Andrews (1991)





Estimating and Predicting Survey Item Quality

Estimating the Quality of Survey Items

Multitrait-Multimethod (MTMM) Model

- Disadvantages for respondents
 - Memory effects 1)
 - Cognitive burden
- Disadvantages for researchers
 - Not possible to repeat all questions 1)
 - Longer questionnaire 2)
 - Time consuming 3)
 - More expensive 4)



Split-Ballot MTMM Model (limits repetitions for respondents)

Survey Quality Predictor (SQP)

SQP from the Program's Perspective



Collection of all existing knowledge about the quality of survey questions ~ 600 MTMM experiments; 6,074 questions; 28 languages; 33 countries



Analysis of experiments and coding of formal characteristics *e.g., domain, balance of request, response scale*



"Meta" analysis of experiments



Implementation of prediction algorithm in software



SQP from the User's Perspective

Coding of the formal item characteristics e.g., number of scale points, polarity

Input



Based on the analysis of many multitrait-multimethod (MTMM) experiments with attitudinal questions



Quality prediction = reliability x validity

Output

66



"Many roads lead to Rome."

Survey Quality Predictor 3.0





A dictionary for questionnaire development



An evaluation tool

comparing qualities, correcting for measurement error



A translation check tool comparing of codings



https://sqp.gesis.org/

SQP as a "Dictionary" for Survey Questions



SQP as **Dictionary**





Database

Filter by 🚯

Study	~	Question owner	~	Language	~
Advanced filters					
By text 🚯				By cod	ing character
Text				Selec	ct characteris

Applied filters				
Studies:	Question owner:	Language:	Country:	Qu

SQP is a searchable database.





SQP as **Dictionary**





Filter by 🕕					
Study	~	Question	owner 🗸	Language	~
Advanced filters					
By text 🚯					By coding character
satisfied					Select characteris
					× RANGE OF THE
					All None
Applied filters					🛃 Bipolar
Studies:	Question	owner:	Language:	Country:	🔲 Unipolar





SQP as **Dictionary**





Database

Filter by 🚯				
Study	~	Question owner	Language	~
Advanced filters				
Applied filters				
Studies:	Question owner:	Language: English	Country:	



72

 $\bullet \bullet \bullet$



• • • • • • • • •
SQP as **Dictionary**

Legend of quality: My codings A Authorized codings O Other users' codings

Que	estion	Study	Language	Country
\uparrow	HS9 / testc9 / <i>Political satisfaction: functioning of democracy</i>	ESS Round 4	English	United Kingdor
Re Ar	quest for answer text: I d on the whole, how satisfied are you with the way d	emocracy works in (Brita	ain/the UK)? Use the same card.	Information

Answer options:	A Authorize
• 0 Dissatisfied	My quality
•2	
• 3	
• 5	
• 6	
•7	

- 8
- . 9
- 10 Satisfied

Filtering is based on all codings. *i.e., my codings, authorized codings, and other users' codings*



Quality		Actions
n 🗛	Replicate Edit	
	Quality	Options
ed prediction	0.712	
y prediction		P

SQP as an **Evaluation Tool**

$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$
$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$
$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$



Consider This Version

Question	Study	Language
A9 / PplFair / social trust, take advantage	ESS Round 4	English

Request for answer text:

Using this card, do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?

Answer options:

- 00 Most people would try to take advantage of me
- 01
- 02
- · 03
- 04
- 05
- 06
- .07
- 08
- 09
- 10 Most people would try to be fair



75



Request for answer text:

Using this card, do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?

Answer options:

- 00 Most people would try to take advantage of me
- 01
- · 02
- · 03
- .04
- .05
- 06
- · 07
- 08
- .09

10 Most people would try to be fair





Or This Version

	Study	Language
estc5 / <i>Social trust:</i> Ivantage	ESS Round 4	English

Request for answer text:

Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? Choose your answer from this card.

• 1 Most people would try to take advantage of me • 2 Most people would try to be fair

And Their Quality Predictions

Quality prediction	Prediction of Question #1	Prediction of Ques	tion #2
Reliability = 1 - random error (r ²)	0.694	0.661	
Validity = 1 - method effect (v ²)	0.836	0.794	
Quality = reliability (r ²) x validity (v ²)	0.580	0.525	
Reliability Coefficient (r)	0.833	0.813	
Validity Coefficient (v)	0.914	0.891	
Quality Coefficient (q)	0.762	0.724	
Method Effect Coefficient (µ)	0.405	0.454	
Reliability Coefficient Interquartile Range	(0.790, 0.890)	(0.770, 0.880)	
Validity Coefficient Interquartile Rang	ge (0.900, 0.960)	(0.860, 0.950)	



SQP as Translation Check Tool

$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$
$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$
$\bullet \bullet \bullet$	$\bullet \bullet \bullet \bullet \bullet$



SQP as Translation Check Tool

- Aim: keeping comparability across languages and countries as high as possible
- SQP: comparison of formal characteristics between translated item and source item (or different countries)
 - Detection of formal differences due to translation (or country adaptation)
 - Allows to suggest corrections
 - But: differences might be due to the properties of the language (or societal context) itself



SQP as **Translation** Check Tool

Compare code

ESS Round 5 / I15 / testd15 / Trust in police, effectiveness quickly arrive / United Kingdom / English VS.ESS Round 5 / I15 / testd15 / Trust in police, effectiveness quickly arrive / Spain / Catalan

Questions	Question #1	Question #2
Study	ESS Round 5	ESS Round 5
Question name	I15 / testd15 / Trust in police, effectiveness quickly arrive	I15 / testd15 / Trust in police, effectiveness quickly arrive
Country/Country prediction	United Kingdom	Spain
Language	English	Catalan
Request for an answer text	If a violent crime were to occur near to where you live and the police were called, how slowly or quickly do you think they would arrive at the scene? Choose your answer from this card, where 0 means very slowly and 4 means very quickly.	Si succeís un delicte amb violència a prop d'on viu vostè i truquessin a la policia, com de lent o ràpid creu que arribaria al lloc del delicte? Triï la resposta d'aquesta targeta.
Answer options	Very slowly Rather slowly Neither slowly nor quickly Rather quickly Very quickly	Molt lent Bastant lent Ni lent ni ràpid Bastant ràpid Molt ràpid
User	authorized coding	authorized coding



download as CSV file

SQP as **Translation** Check Tool

Codings	Coding of Question #1	Coding of Question #2
> Domain	National politics	National politics
> Domain: national politics	Local institutions	Local institutions
> Concept	All other simple concepts	All other simple concepts
> Concept: other simple concepts	Evaluation	Evaluation
> Social Desirability	Not present	A bit
> Centrality	Not at all central/salient	A bit central
> Reference period	Present	Future
> Formulation of the request for an answer: basic choice	Direct request	Indirect request
> WH word used in the request	WH word used	WH word used
> 'WH' word	How (extremity)	How (intensity)
> Request for an answer type	Interrogative	Imperative
> Use of gradation	Gradation used	Gradation used
> Balance of the request	Balanced or not applicable	Balanced or not applicable
> Presence of encouragement to answer	No particular encouragement present	Encouragement present
> Emphasis on subjective opinion in request	Emphasis on opinion present	Emphasis on opinion present



SQP as **Translation** Check Tool

Quality prediction	Prediction of Question #1	Prediction of Question #2
Reliability = 1 - random error (r ²)	0.670	0.673
Validity = 1 - method effect (v ²)	0.819	0.728
Quality = reliability (r ²) x validity (v ²)	0.549	0.489
Reliability Coefficient (r)	0.818	0.820
Validity Coefficient (v)	0.905	0.853
Quality Coefficient (q)	0.741	0.700
Method Effect Coefficient (µ)	0.425	0.522
Reliability Coefficient Interquartile Range	(0.790, 0.870)	(0.800, 0.880)
Validity Coefficient Interquartile Range	(0.870, 0.960)	(0.800, 0.950)



 $\bullet \bullet \bullet$

 $\bullet \bullet \bullet$

.

Limitations

- Only for survey questions measuring continuous latent variables ≠ factual, observable questions such as demographics, facts, behavioral questions
- Only for formal characteristics cannot substitute cognitive pretesting, expert review, or web probing techniques
- Empirical program aimed at prediction \neq explanation
- Prediction limited to type of questions in the MTMM experiments







Questionnaire Translation as a Special Type of Translation

- Comparability/equivalence to the source
 questionnaire in terms of meaning and measurement
 characteristics (e.g., scales or other design features)
- Adherence to the grammatical, syntactical, and idiomatic requirements of the target language
- Adherence to general questionnaire requirements (e.g., clarity, conciseness, fluency, adequacy for mode, wording conventions)





Translation in More Detail

- Exceptions and trade-off decisions may be necessary
- Translation is always an individual, case-by-case decision
- Translation is always dependent on specific language combinations and the specific questionnaire context
- A Next slides for raising awareness of potential challenges







Meaning is not static.

Meaning is determined by context (including co-text).

Meaning is determined by the person. (See pretesting and different associations coming up.)





Meaning – Problems I

Obvious misinterpretations

• "X tells me how I'm doing" translated as "X m'aide/ me corrige" [helps me/corrects me] (Candell & Hulin, 1986)





Meaning – Problems II

Shifts in meaning that affect measurement

- E.g., formulation is too narrow or too wide
- "If people who have come to live here commit any crime, they should be made to leave." (ESS R1) (Davidov et al., 2014; Fitzgerald et al., 2011)
- The chosen Danish translation ("lovovertrædelse" for crime) roughly translates back as 'breach of the law' – this translation is quite vague and general and thus had a negative impact on comparative measurement



Meaning – Problems III

Omission of meaningful text elements

- Text elements that specify a temporal, spatial, or otherwise conceptual frame within which to answer
- E.g., 'in general', 'on average', 'during the past two weeks', 'rarely', 'often'



Knowledge of questionnaire design principles important

Meaning – Problems IV

Unsuitable connotations

- The item "He looks for adventures" was literally translated into French as "Il cherche les aventures"
- But: the translation carries sexual connotations that do not fit to the measurement intention
- Solution: change plural 'adventures' into singular 'adventure'
 - "Il recherche l'aventure" (Supplementary questionnaire, e.g., ESS round 6)



Meaning – Problems V

Unclear Meaning

Ambiguity or potential misunderstanding

Unnecessary addition of text elements

Note. Translation sometimes longer or more explicit because of linguistic reasons that cannot be avoided.



Response Scales

Important aspects

- Dimension (satisfaction, ability, agreement, etc.)
- Type of polarity (unipolar vs. bipolar)
- Quantification or negation (very, quite, not, dis-, etc.)
- Symmetry
- Relationship or distance between scale points
- Balancing
- Yes-no vs. more open-worded questions (To what extent...?)





Response Scales – Polarity and Extremeness

- Goal: producing a bipolar scale (good bad)
 [unipolar scale: good not good]
- Goal: translation should match the *extremeness* (endpoints)
- Depending on language words like 'extremely', 'completely', 'fully', etc.

B24	CARD 1 of educa	0 Now, u tion ¹⁵ in [sing <u>this</u> [country]	card, ple nowaday	ase say /s?	wha
Extre bad	mely					
00	01	02	03	04	05	(





Response Scales – Quantification: Symmetry

- Goal: symmetrical set-up
- very, somewhat ... somewhat, very

B4 PolDcs. **CARD 8** How difficult or easy do you find it to make your mind up⁷ about political issues⁸? Please use this card.



Very	difficult	1
------	-----------	---

- Difficult 2
- Neither difficult nor easy
 - Easy 4

3

5

8

- Very easy
- (Don't know)

Response Scales – Relationship: Disjunct Values

- Goal: categories should remain disjunct in translation
- Disjunct = excluding each other
- Especially relevant for numerical scale points

A1	TvTot CARD 1 On an average weekday, how much time, in total, do you spend watching television? Please use this card to answer.
	No time at all
	Less than ½ hour
	¹ / ₂ hour to 1 hour
	More than 1 hour, up to1½ hours
	More than 1½ hours, up to 2 hours
	More than 2 hours, up to 2½ hours
	More than 2½ hours, up to 3 hours
	More than 3 hours
	(Don't know)





A Note on Scale "Translation"

 Still an open field for research – a lot of things change in questionnaire scale "translation" (impact often unknown)

by linguistic necessities (e.g., a bipolar/symmetrical scale becomes a unipolar/asymmetrical scale)

by intentional decisions (e.g., using home-grown scales rather than comparable scales)

by cultural needs

(e.g., taking into account communication and discourse norms)

or *inadvertently*



Scale "Translation" – Example



Behr & Shishio, 2016





Gender-Inclusive Language

- Correct way of addressing people and inclusive language is increasingly important
- Many things are still in evolution
- Different languages have different needs when it comes to
 - ... accommodating gender, honorific titles, etc.; with potential impact on survey wording and design (slashes/brackets/genderspecific items or questionnaires; accommodation of the third gender)
 - ... taking into account societal developments



Checking of Linguistic Correctness





100

And What About Adaptations?

Think about a version for your country. What could a translation or adaptation look like?

- 1) Are you able to walk a block on flat ground?
- 2) Are you able to push open a door after turning the knob?
- 3) Dress like the way local people do
- 4) Follows local media
- 5) Able to eat local food







Some Concluding Remarks on Translation

- Sometimes several good translations exist
 - There is no such as thing as "one translation and everything else is wrong"
- Sometimes optimal translations exist
- Sometimes only suboptimal translations are possible
- Translation is a (trade-off) decision-making process, influenced by equivalence needs, target language requirements, questionnaire design principles, and conversational needs – and driven by the translation brief...





Translation as a Decision-Making Process

Brief – key variables of the translation

purpose





Individual translation & strategies



Translation Brief (Specifications)

All translation starts with a translation brief

Information on the study and the questionnaire

Information on the target group (age ['you'], education level, language, any other particularities)

Survey mode (web, face-to-face, phone, etc.)

Specific requirements for the translation (tone or style, adherence to a glossary, consistency with other questionnaires, leeway for adaptation, etc.)



False Friends

- Be aware of words in your language sounding similar to English [source] language words
- Depending on language pairs
- **Examples:** intimate single simple card individual control – handy – job
- Good knowledge of English language (source language) needed
- Use of monolingual dictionaries helpful





Grammar and Syntax I

- Don't stick too closely to English (source) grammar or syntax if the target language is structured differently
- The less one is experienced in translation, the more one translates on the word level
- Find a good balance between faithfulness and fluency





Grammar and Syntax II

- Examples
 - Information is a singular noun in English, but in other languages it is often a plural noun
 - leave it out completely? \rightarrow always a case-by-case decision needs to be translated in a way that fits the survey style in the
 - "if any"/"if at all" in other languages this often get longer term • "Now I'd like you to tell me your views on various issues." \rightarrow target country



Supposed One-To-One "Equivalencies"

- A single word in English $\leftarrow \rightarrow$ several possible translations (e.g., administration, environment)
- Translation depends on context
 - *Government:* "How successful is the government nowadays in controlling crime?" state – governing body/people – public administration?
 - You: Singular or plural? Polite form? General form?
 - *Fairly:* equally or just or rather?


Common Translation Mistakes





How to Reduce These Risks?

Best Practice

 TRAPD = translation, review, adjudication, pretesting, documentation



- Key elements
 - Multi-step
 - Multi-disciplinary
 - Cooperation within a team





Mohler, Dorer, de Jong, & Hu (2016)

Translating Staff

- Translators: skilled translation practitioners, translation into mother tongue, experience in/briefing on questionnaire translation
- **Reviewer(s)**: besides translation competence, knowledge of the study, the topic, and questionnaire design
- Adjudicator(s): besides language competence, knowledge of the study and the topic





Cross-Cultural Questionnaire Development

- Deliberately designed cross-cultural surveys pave the way for translation/anticipated adaptations
- The source questionnaire is constructed taking both translation and adaptation into account, usually the focus is on translation, though
- The following procedures apply to source questionnaires that can be translated but also those that have some form of anticipated adaptation



112

Interplay Between Q. Design and Translation/Adaptation

- Good and comparable translation/adaptation presupposes adequate source questionnaire design
 - → consideration of cultural and linguistic aspects already during the design of the source questionnaire
- Many problems that otherwise would come up during translation or even during data analysis can thus be prevented





Development of Source Questionnaires

Differences in Intercultural Feedback

- Sequential: more or less monocultural development of source questionnaire, then translation
- Parallel: intercultural input during the development of the source questionnaire, then translation/adaptation
- Simultaneous: simultaneous development of the questionnaire in several languages











A Bag of Tips

- If you plan an international study, collaborate with partners from the different countries/cultures
- Have a draft questionnaire/translation undergo different types of checks – the more, the better
- Plan ahead: Develop a plan when and where to integrate various feedback/checks



116

ergo etter re to



Questionnaire Design for Cross-Cultural Surveys

Iydia.repke@gesis.org
https://www.linkedin.com/in/lydiarepke/
@lydiarepke.bsky.social
@LydRep