Prediction of the quality of survey questions using the program SQP: background and applications

Lausanne, 1st November 2016

Melanie Revilla, RECSM-UPF

melanie.revilla@upf.edu

No measurement without errors

Most social sciences analyses assumed that themeasurement is perfectNot true!

Measurement errors (ME) quite large

On average, **≈50% of the variance** of the observed variables is due to error (Alwin 2007)

But different questions have **different sizes** of ME

Measurement error and quality

Quality = part of variance explained by the latent concept of interest

 \rightarrow complement of measurement errors



Measurement error and quality

Quality = part of variance explained by the latent concept of interest

 \rightarrow complement of measurement errors



Level of quality

Quality = standardized coefficient which varies from 0 to 1; the closer to 1, the better

In practice:

- Always some measurement errors, thus quality always lower than 1
- Typical to get quality estimates between .50 and .80

The quality varies depending on:

- The topic
- The question formulation
- The response scale
- The mode of data collection
- The country
- The language
- Etc.

How to evaluate the quality?

2 main alternatives

- 1. Estimation using the **Multitrait-Multimethod** (MTMM) approach
- 2. Prediction using the **Survey Quality Predictor** (SQP) software

Repeat several questions using different methods

Example (ESS round 1)

- 3 questions (each measuring one trait):
 - On the whole, how satisfied are you with the present state of the economy in [country]?
 - Now thinking about the government, how satisfied are you with the way it is doing its job?
 - And on the whole, how satisfied are you with the way democracy works in [country]?
- Repeated with 3 methods:
 - M1 = 11 pts (extremely)
 - M2 = 4 pts (very)
 - M3 = 6 pts (extremely)

Origins of the approach: Campbell & Fiske (1959)

They study discriminant validity, looking directly at the MTMM matrix

		Method 1			M	lethod	2	Method 3		
		trait ₁	trait ₂	trait ₃	trait ₁	trait ₂	trait ₃	trait ₁	trait ₂	trait ₃
	trait. /	1		1				1		
Method ₁	traita	.6	1)
-	trait	.7	.8	1						
	trait ₁	.5	.3	.4	1					
Method ₂	trait ₂	.3	6	4	.8	1				
-	trait ₃	.2	.3		.6	.7	1			
	trait ₁	.7	.3	.2	.6	.4	.3	1		
Method ₃	trait ₂	.2	.5	.1	.3	.6	.2	.7	1	
	trait3	.4	.2	.6	.2	.1	.7	.9	.8	1 /

The true score model: Saris & Andrews (1991)

Analyses using SEM Equations of the model: $\begin{cases} Y_{ij} = r_{ij} T_{ij} + e_{ij} \\ T_{ij} = v_{ij} F_i + m_{ij} M_j \end{cases}$



Split-ballot MTMM: Saris, Satorra & Coenders (2004)

Limit of the traditional approach

- 3 repetitions needed for identification \rightarrow Burden & Memory effect

Split Ballot-MTMM:

- Combine advantages of the MTMM approach with the ones of the SB approach in order to reduce the number of repetitions
- Respondents randomly assigned to different groups, each group gets only 2 methods

Grouns	Time 1	Time 2	Groups	Time 1	Time 2
1	Mothod 1	Mothod 2	1	Method 1	Method 2
1	Method 1	Method 2	2	Method 2	Method 3
Ζ	Method 1	Method 3	3	Method 3	Method 1

- Model still identified under quite general conditions

Limits of the MTMM approach

Limits of the 2-group split ballot design

Saris et al. (2004) warned for problems with the 2-group design:

- 1. When the correlations between the traits are very close to zero
- 2. When the correlations between the traits are identical

They saw this as very special cases, but in practice the problems are common (Revilla & Saris 2013)

Experiments	NC	HC	PS	Total
R1 - Media use	15	4	0	19
R1 - Pol. efficacy	1	11	7	19
R1 - Pol. orientation	4	8	7	19
R1 - Satisfaction	3	9	7	19
R1 -Social trust	3	13	3	19
R1 - Political trust	2	10	7	19
R4 - Media use	16	6	0	22
R4 - Satisfaction	9	10	3	22
R4 - Political trust	1	13	8	22
Total	54	84	42	180
In percentages	30%	46.7%	23.3%	100%



NC : non-convergence HC: Heywood cases PS: proper solution

Limits of the MTMM approach

Limits of all MTMM experiments

Get information only after data collection

Not adapted for background variables

- Often difficult to think of alternative formulations
- Evaluation of such questions can be done with Quasi Simplex models

Results for specific questions

- Cannot be generalized to other measures
- Many interactions

Repetitions needed so require long surveys

- Time, Cost, Cognitive Burden, Memory effect

In practice, impossible to repeat all questions

- Alternative? The SQP software



What is SQP?

A survey quality prediction system for questions used in survey research and a database of questions with information about their quality

What can you get in SQP?

- Reliability, validity, and quality coefficients
- With confidence intervals

What do you have to do?

- Go to <u>sqp.upf.edu</u>



- Consult the information already stored in the database
- Introduce a new question and code its characteristics following the program instructions until you get a prediction

What is behind the program?

25 years of cumulative experience



A meta-analysis of 3,700 quality estimates from MTMM experiments conducted in \approx 30 countries (mainly ESS)

These estimates are explained by up to 60 questions characteristics

Random Forest approach used

Good predictions obtained: explained variance (R²) for reliability of .65 and for validity of .84



Limits of SQP

Limits of the current version



In this section the current limits of SQP software version will be summarized.

SQP was developed using the data from more than 250 Multitrait – Multimethod (MTMM) experiments ca quality was known. Based on this database of questions, the MTMM quality estimates could be related to and Gallhofer, 2014). The kind of survey questions for which SQP can provide a reliable prediction, is lim

The SQP team expects most of these limits to be improved in a new version of SQP. In the meantime, ple



All this can be improved in the next version

Limits of SQP

More general limits

The coding of some characteristics is somehow subjective

- Example: centrality or social desirability

There is an error in the estimation of the size of the error

- Important to keep this in mind
- But difficult to take it into account in the analyses

The predictions tend to be more "in the middle"

- Predictions based on thousands of estimates
- If you study a deviant case, this may not be well predicted

More information





RECSM working paper:

Saris, Oberski, Revilla, Zavala, Lilleoja, Gallhofer & Gruner (2011). "Final report about the project JRA3 as part of ESS Infrastructure".

Tutorials: <u>https://www.youtube.com/channel/UCpljiQFlE4j5CYI-rqMKDig</u>

What to do with once we know the quality?

Different uses of the information about quality

<u>Before</u> the data collection:

- Compare different formulations of the questions and use this information to improve questionnaire design
- Compute the quality by groups (countries/languages) to see we can compare standardized relationships across groups
- Compare the codes of the questions in different languages to detect deviations in translation

<u>After</u> data collection:

Correct for measurement errors

Questionnaire design

Example: 2 questions to measure "importance of the value honesty" (Revilla, Zavala Rojas & Saris, 2016)

Q3a- To what extent do you think that honesty is important?							Which one is better? Some indications in the literature
Not				Very	Don't Know		- About AD scales versus IS
important 1	2	2	4	important 5			- About fixed reference points
1	2	3	4	2			

But ...

because of the huge number of decisions and the potential **interactions** between them, one should estimate the **quality** of both questions to know which form it is better to use Q3b-Now we want to ask you questions about different values. Do you agree or disagree with the following statement: "honesty is important". Please use this card to answer.

- 1 Agree strongly
- 2 Agree
- 3 Neither agree nor disagree
- 4 Disagree
- 5 Disagree strongly
- Don't Know

Comparison across groups

Conditions for comparing standardized relationships

If the quality varies across groups

And it usually does

- Observed correlations can be the same in several groups whereas the true correlations are different
- Observed correlations can be different in several groups whereas the true correlations are the same

Correlations, regression coefficients

Countries or time points

Standardized estimates can only be compared across groups if the quality of the questions is the same in the different groups.

Example: ESS round 3 data

Correlation between "love learning new things" (11 AD scale) and "feel what I do is valuable" (7 AD scale)



Detection of deviations in translation

Example of the ESS

Translation procedure in the ESS

- 2 persons translate independently from English to a given language
- A third person compares both and decides on the final translation
- Still unnecessary deviations in the formal characteristics

New step added to the procedure in more recent rounds

- Questions are coded in SQP
- Codes of the questions are compared
- Unnecessary deviations can be corrected before the fieldwork

Correction for ME: an illustration

Model we want to study



What is the problem? Model not identified!

Most common way to deal with this

Assuming no ME \Leftrightarrow q=1 \Leftrightarrow Model without correction



How to correct for ME?

Fix the loadings to the values of the quality coefficients

Information about quality can be obtained from

- MTMM experiments
- SQP predictions

Big advantage: Does not require to collect previous data Big advantage: Can be obtained for almost any question

Correction for ME using quality coefficients

Step 1: get the predictions of quality for each question using SQP Step 2: fix the loadings to these values



Estimates of the effects

Rather different conclusions with respect to the size of the effects

Effects	No correction	fix ly _{sQP}
Better to Allow	23	17
Eco to Allow	31	56
Eco to Better	.45	.52
Cult to Better	.41	.40
Corr Eco-Cult	.66	.82
Expl. var Allow	.25	.50
Expl. var Better	.61	.78

More information

WILEY

Design, Evaluation, and Analysis of Questionnaires for Survey Research



Willem E. Saris Irmtraud N. Gallhofer

WWW.

WILEY SERIES IN SURVEY METHODOLOGY

Conjugation winter of

ESS EduNet Countries by Round

About

Topics

Measurement errors Chapter 1 Chapter 2 Chapter 3 Chapter 4 Chapter 5 Chapter 5 Chapter 7 Appendix 1 Appendix 1 Appendix 2 References Multilevel models

A simple procedure to correct for measurement errors in survey research

By Anna DeCastellarnau and Willem Saris

ESS EduNet European Social Survey Education Net

Introduction

Although most applied researchers believe that survey data contain measurement errors, very few correct these errors. In principle, the reason for this omission cannot be that the procedures are not known because they were already developed in the 20th century [Gol73]. One of the reasons for this omission is probably that these procedures make it necessary to collect multiple indicators for all variables in the study in order to correct for measurement errors. This doubles or triples the response burden for the respondents, increases the costs of the research and makes the analysis rather complex. The purpose of this text is therefore to illustrate how corrections can be made in a very simple way and to show that researchers can and should always correct for these errors.

http://essedunet.nsd.uib.no/cms/topics/measurement/

Saris, W.E., & M. Revilla (2016). "Correction for measurement errors in survey research: necessary and possible". *Social Indicators Research*, 127(3): 1005-1020. DOI: 10.1007/s11205-015-1002-x

Conclusions

10 main points to remember

- 1. There is no measurement without errors
- 2. These measurement errors affect the results and comparisons across groups
- 3. It is crucial to evaluate the size of these errors
- 4. This can be done using an MTMM approach
- 5. But the MTMM approach has limits
- 6. An alternative is to use the SQP software
- 7. Allows to predict the quality of the questions before going to the field
- 8. Can be used to improve questionnaire design and avoid deviations in translation
- 9. Still, there will be measurement errors so it is necessary to correct for this
- 10. SQP has its own limitations and further research is needed to improve it and increase its scope (e.g. to more languages or web-specific scales)

Thank you for your attention!



Questions of the illustration

Form of the questions in the Main questionnaire of the ESS Country: Ireland

First question of the battery (not used)

To what extent do you think Ireland should allow people of the same race or ethnic group as most Irish people to come and live here?

Opinion about immigration (Allow)

How about people from the poorer countries outside Europe? Use the same card

- 1. Allow many to come and live here
- 2. Allow some
- 3. Allow a few
- 4. Allow none

Opinion about economic consequences (Eco)

Would you say it is generally bad or good for Ireland's economy that people come to live here from other countries? Please use this card.

Bad for the										Good for the	
economy										economy	
0	1	2	3	4	5	6	7	8	9	10	

Opinion about the cultural consequences (Cult)

And, using this card, would you say that Ireland's cultural life is generally undermined or enriched by people coming to live here from other countries?

Undermined

Enriched

Consequences for the life in general (Better)

Is Ireland made a worse or a better place to live by people coming to live here from other countries? Please use this card.

Worse										Bette
place										place
to live										to live
0	1	2	3	4	5	6	7	8	9	10