

Split questionnaire design: can a question battery be split and still produce the same measurements? Evaluating context effects in the study of moral values*

Jimena Sobrino Piazza

*Thesis for obtaining a MA in Public Opinion & Survey Methodology (supervised by Prof. Caroline Roberts)

Introduction

- In a questionnaire, earlier questions may impact answers given to later questions.¹
- In a moment in which the split questionnaire design (questionnaire modularization) becomes increasingly popular, the aim of this study is to contribute in the refining of these splitting strategies, in order to minimize the appearance of context effects.
- A particular question setting is studied: question batteries in which items related to different constructs are intermixed and share a common rating scale. Can a question battery be split and still produce the same measurements?
- The idea is to shed light on the possible context effects that may be introduced if the question battery items are changed, both at the item- and multi-item level of analysis.

Split questionnaire design (SQD)



- The SQD consists in splitting the original questionnaire into modules, and to combine them, to create different, unique, shorter versions of the questionnaire, which are then administered to different subsamples.

Question batteries

- The particularity of this question setting is that all items share a common rating scale. Previous items may act as anchors or reference points for the use of the rating scale.
- If some of the items in the battery measure different multi-item constructs, splitting the QB could introduce measurement error on both item and scale level.

Moderators of context effects: attitude strength

Generalized assumption that stronger attitudes should be associated with more chronically accessible beliefs. Thus, given that chronically accessible beliefs are context independent, stronger attitudes should be less affected by context.²

Research questions:

- What are the effects of splitting a question battery on measurements at the item level?
- To what extent are any observed effects of splitting at the item level moderated by respondent characteristics?
- To what extent do effects of splitting a battery at the item level affect multi-item measures embedded in the battery?

Methods

Data

Data was taken from the Swiss European Values Study (EVS) 2017, which had implemented an experimental design to compare the feasibility of online surveys, both on a SQD and in full-length design.

Question battery on moral beliefs

A particular question battery was studied. Given the SQD, the question battery of moral beliefs was administered in two versions: one containing all items together (whole version) and another presenting items in two short batteries, separated thematically (split version), according to the two moral dimensions embedded in the battery, self-determination and civic morality.

Please tell me for each of the following whether you think it can always be justified, never be justified, or something in between, using this card.

Whole version

- Claiming state benefits
- Cheating on tax
- Taking soft drugs
- Accepting a bribe
- Homosexuality
- Abortion
- Divorce
- Euthanasia
- Suicide
- Having casual sex
- Avoiding a fare
- Prostitution
- Artificial insemination
- Political Violence
- Death penalty

Split version

Self-determination morality

Module 1

- Taking soft drugs
- Homosexuality
- Abortion
- Divorce
- Suicide
- Having casual sex
- Prostitution
- Artificial insemination

Civic morality

Module 2

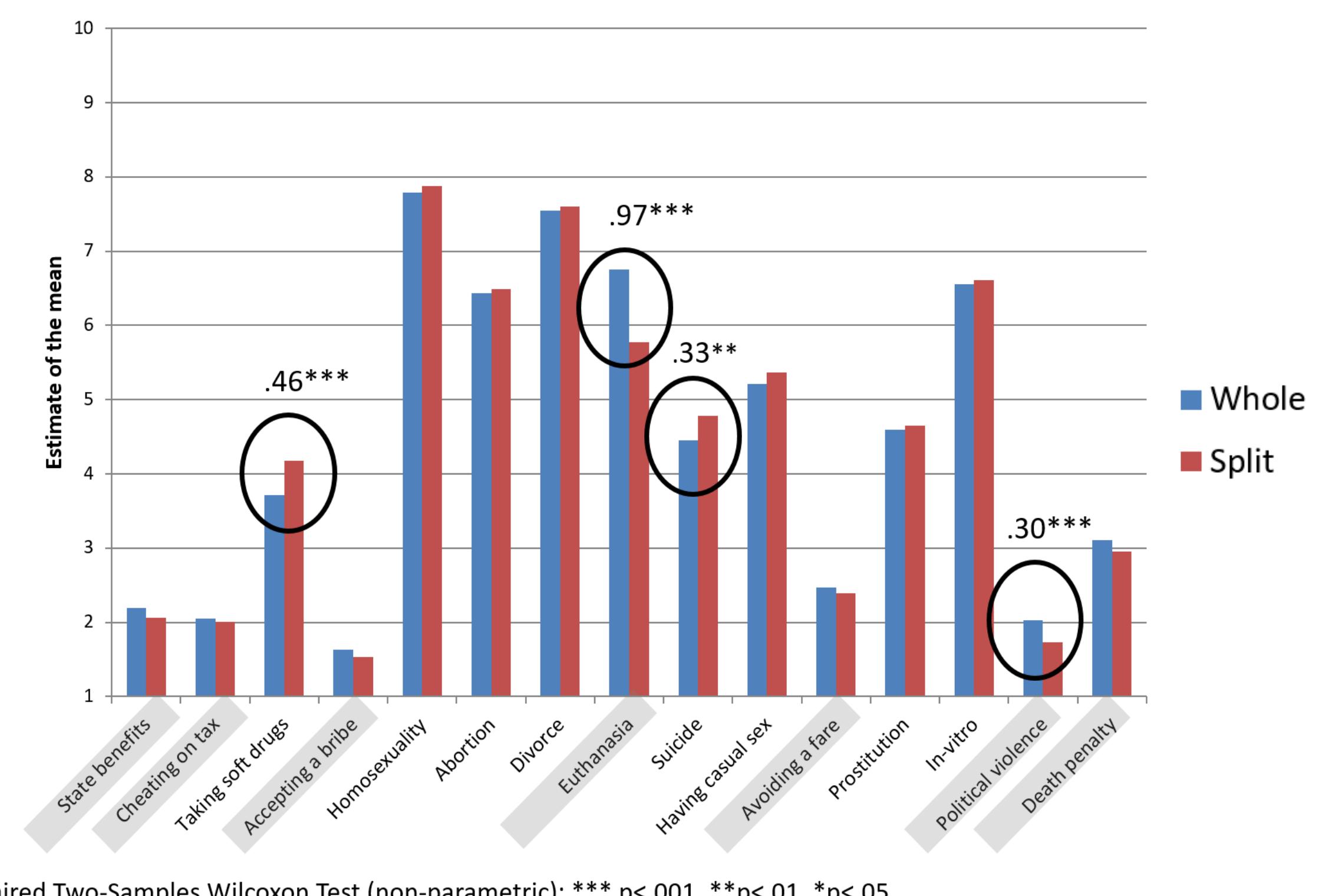
- Claiming state benefits
- Cheating on tax
- Accepting a bribe
- Euthanasia
- Avoiding a fare
- Political Violence
- Death penalty

Analytical approach

- The measurements produced by the two groups (whole & split) were compared.
- In a preliminary step, the sample composition of the two groups was compared, in order to control for bias.
- RQ1: Effects at the item level:** The estimates of the mean were compared. The non-parametric unpaired Two-Samples Wilcoxon Test was used to assess the statistical significance of the differences.
- RQ2: Effects moderated by respondent characteristics:** Subgroup comparison of the mean, according to 4 explanatory variables of moral beliefs: religiousness, age, education, political orientation.
- RQ3: Effects at the scale level:** Reliability analysis & Multigroup Confirmatory Factor Analysis.

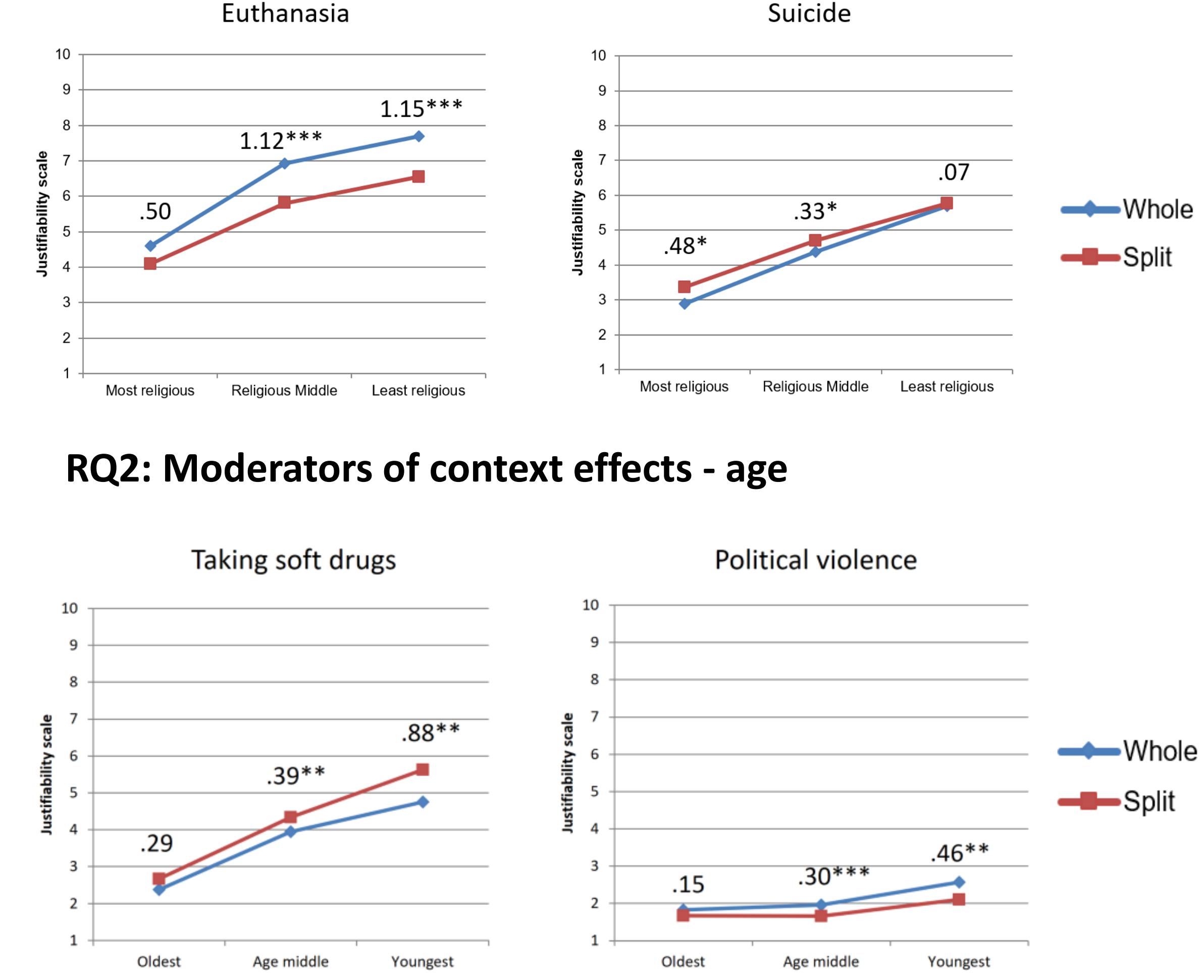
Results

RQ1: Effects at the item level – comparison of the means

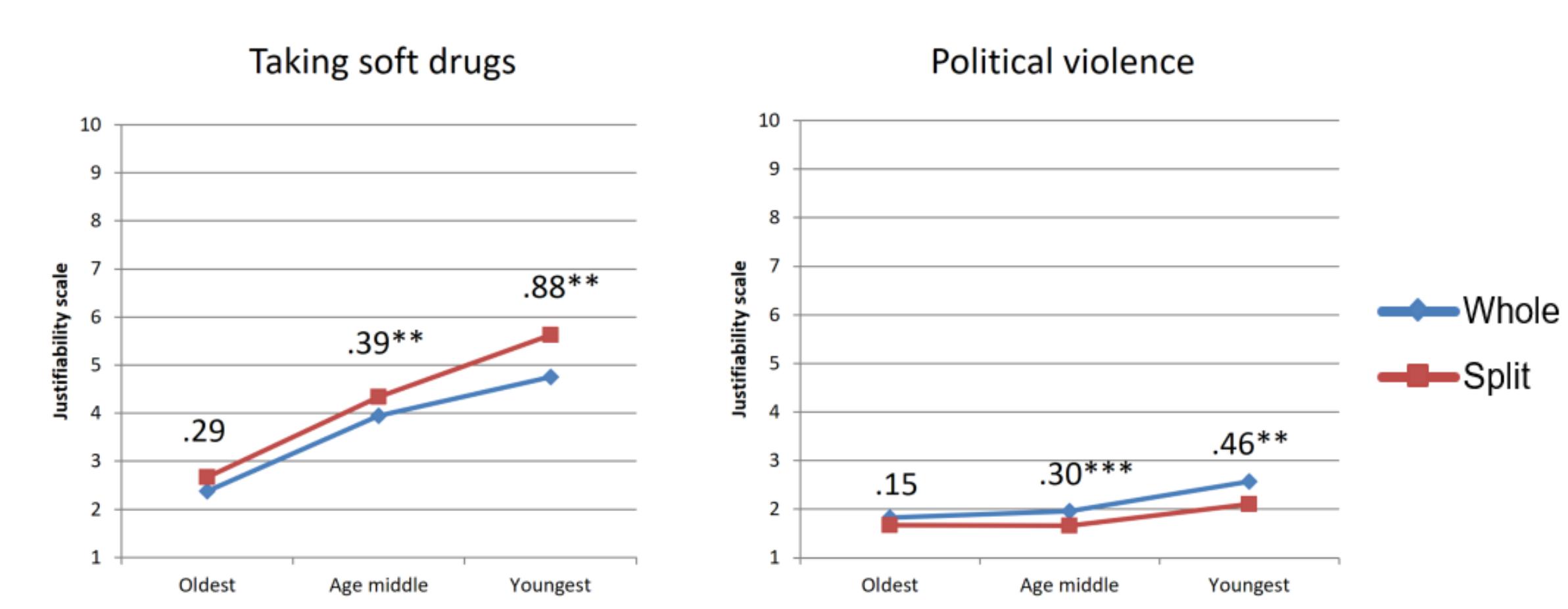


Unpaired Two-Samples Wilcoxon Test (non-parametric): *** p<.001, **p<.01, *p<.05.

RQ2: Moderators of context effects - religiousness



RQ2: Moderators of context effects - age



RQ3: Effects at the scale level – reliability analysis

	Whole		Split	
	Cronbach's Alpha	N (% valid)	Cronbach's Alpha	N (% valid)
Self-determination morality	.835	956 (94.7)	.813	1472 (94.5)
<i>If item deleted:</i>				
Taking soft drugs	.824		.797	
Homosexuality	.820		.791	
Abortion	.805		.779	
Divorce	.808		.782	
Euthanasia	.826		.825	
Suicide	.817		.790	
Having casual sex	.806		.777	
Prostitution	.820		.794	
Artificial insemination	.835		.806	

RQ3: Effects at the scale level – MCFA

Latent concept	Final model	χ^2	df	CFI ^a	RMSEA ^a	SRMR ^a
Self-determination morality	Partial Strict, equal scale means (free: intercept of euthanasia)	603.570	79	.913	.074	.051
Civic morality	Partial Strict, equal scale means (free: residuals of political violence & accepting a bribe)	112.183	22	.966	.057	.042

Notes:

^a: CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Analysis and conclusions

- At the item level**, out of the fifteen items embedded in the battery, four presented significantly different means across the two battery versions. The biggest difference was for the item euthanasia, which was allocated away from the other items of the self-determination morality.
- The item of suicide seemed to be affected by the presence or absence of the item of euthanasia, suggesting the existence of a part-whole combination¹ between those items.
- The subgroup comparison** showed that the effect in the item suicide was stronger among the most religious people. Whereas the literature assumes strong attitudes as a moderator of effects of context, here the effect could be introduced precisely due to a strong attitude towards life and death.
- At the scale level**, context effects did not affect scale means. The thematic split did not increase artificially constructs reliability.
- The effect of euthanasia was strong enough to undermine the reliability, as well as the scalar invariance of the self-determination morality construct.
- Thematic split of a QB does not seem to introduce important differences in the measurements, unless the thematic allocation of items is done inappropriately.

References

- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic Press.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*, 103(2), 299-314.