PC and mobile web surveys: grid or item-by item format?

Aigul Mavletova, NRU Higher School of Economics, Russia

Mick P. Couper, Survey Research Center, University of Michigan

Daniil Levedev, NRU Higher School of Economics, Russia

Grid format

Отметьте, пожалуйста, насколько Вы доверяете разным категориям людей?

	Совсем не доверяю	Скорее не доверяю	Скорее доверяю	Полностью доверяю
Соседи	✔	✔	✔	✔
Люди, с которыми Вы впервые встретились	✔	✔	✔	✔
Семья	✔	✔	✔	✔
Люди, с которыми Вы лично знакомы	✔	✔	✔	✔
Люди другой национальности	✔	✔	✔	✔
Люди другой религии	✔	✔	✔	✔

Item-by-item format

Отметьте, пожалуйста, насколько Вы доверяете разным категориям людей?

Люди, с которыми Вы впервые встретились

Совсем не доверяю

Скорее не доверяю

Скорее доверяю

Полностью доверяю

Отметьте, пожалуйста, насколько Вы доверяете разным категориям людей?

Соседи

Совсем не доверяю

Скорее не доверяю

Скорее доверяю

Полностью доверяю

Background

PC web

➢ shorter completion times, higher non-differentiation, higher item nonresponse, somewhat lower validity in the grid than in the item-by-item format (Callegaro, Shand-Lubbers, and Dennis, 2009; Couper, Traugott, and Lamias, 2001; Toepoel, Das, and Van Soest, 2009; Tourangeau, Couper, and Conrad, 2004; Peytchev, 2005).

➢ measurement equivalence depends on the number of response options (Liu and Cernat, 2016).

Mobile web

➢ similar Cronbach's alpha coefficients and longer completion times in the grid than in the item-by-item format among mobile web respondents (Revilla, Toninelli, and Ochoa, 2017)

➢ longer completion times in the item-by-item paging design among mobile web respondents than in the grid format among PC web respondents (De Bruijne et al., 2015).

What is the effect of an item-by-item scrolling format relative to grids for both PC and smartphone respondents?

Main Hypothesis

- **Item-by-item format** is associated with **higher data quality and stronger measurement equivalence** than the grid format across devices (smartphones and PCs).

- This effect is **stronger among smartphone respondents**.

Experimental Design

Two-wave experiment with crossover design

Factors

Format (grid vs. item-by-item)

Device (smartphone vs. PC)

Group	1st wave	N	2nd wave	N
1	Grids on smartphone*	258	Grids on PC	165
2	Grids on PC	285	Grids on smartphone*	175
3	Item-by-item on smartphone	278	Item-by-item on PC	193
4	Item-by-item on PC	292	Item-by-item on smartphone	176
5	Item-by-item on smartphone	276	Grids on PC	182
6	Grids on PC	289	Item-by-item on smartphone	188

*Survey was not optimized for smartphones

Portrait mode:

non-optimized

Landscape mode:

Data Collection

	1st wave	2nd wave
Fieldwork	December, 2016	January – February, 2017
Participation rate	3.6%	64.3%
"Wrong" device	57.8%	15.3%
Completes	1678	1079

- Volunteer online access panel (Online Market Intelligence) in Russia, used mobile devices to access the Internet in the previous 30 days
- Tablet respondents excluded (1st wave = 82; 2nd wave = 56)
- Email invitation with a standard incentive

Questionnaire

111 items:

• Main focus – trust

• 7 question sets (49 items), most questions – World Values Survey

All questions were **obligatory**

	Number of items	Latent factors	Response scale
Set 1: Risk willingness	4 items	1 factor	4-point scale
Set 2: Trust	6 items	2 factors	4-point scale
Set 3: Caution	6 items	2 factors	4-point scale
Set 4: Moral and rational trust	6 items	2 factors	7-point scale
Set 5: Institutional trust	8 items	2 factors	5-point scale
Set 6: Tolerance	8 items	3 factors	4-point scale
Set 7: Schwartz values	11 items	2 factors	6-point scale

Indicators

1. Breakoff rates

2. Completion time

3. Concurrent validity

4. Straightlining

5. Measurement equivalence

6. Test-retest reliability

7. Subjective indicators of respondent burden

1. Results: breakoff rates

Condition	Format		Device	
	Grids	Item-by-item	PC	Smartphone
Wave 1	12.4%	12.5%	10.0%***	15.2%***
Wave 2	4.2%	3.8%	3.9%	4.1%

*** $p < 0.001$

No significant format*device interaction.

2. Results: completion time

▶ All differences are statistically significant

▶ Longest completion time: mobile item-by-item

Format		Device		Format x Device			
Grid	Item-by-item	PC	Mobile	PC gr.	PC item	Mob. gr.	Mob.item
18.8	20.9	19.1	20.8	18.9	19.4	18.7	21.7
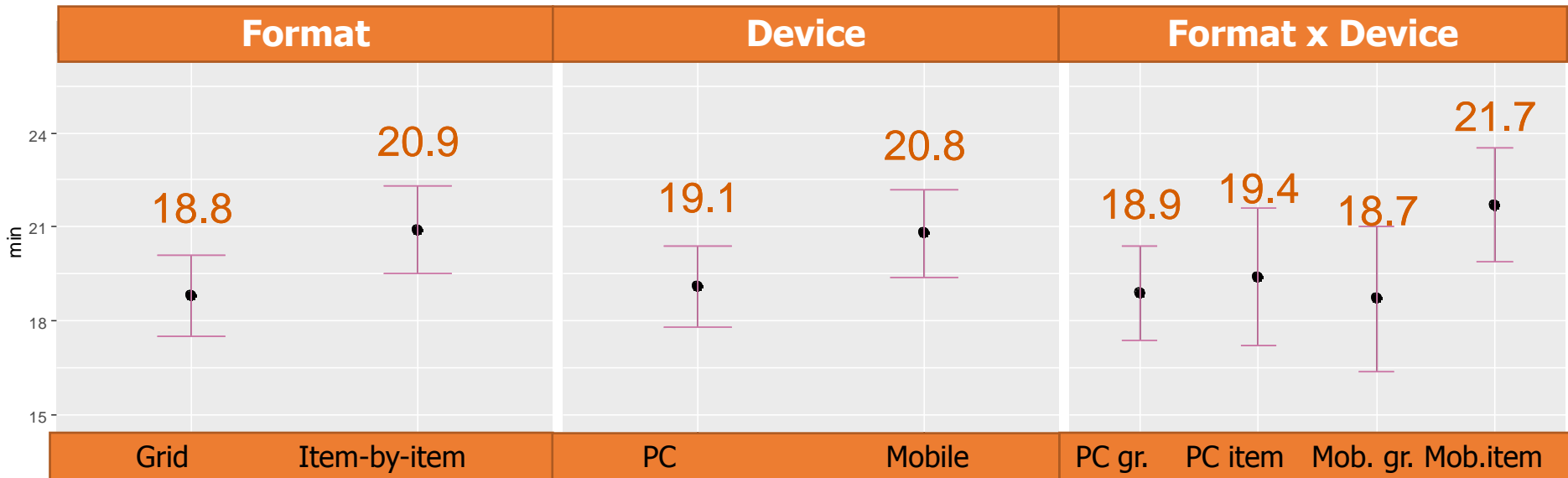
t(df=1676)=4.48***   t(df=1676)3.58***   F(3,1674)=10.26***

Wave1, similar results in Wave 2

3. Results: Concurrent validity (Format)
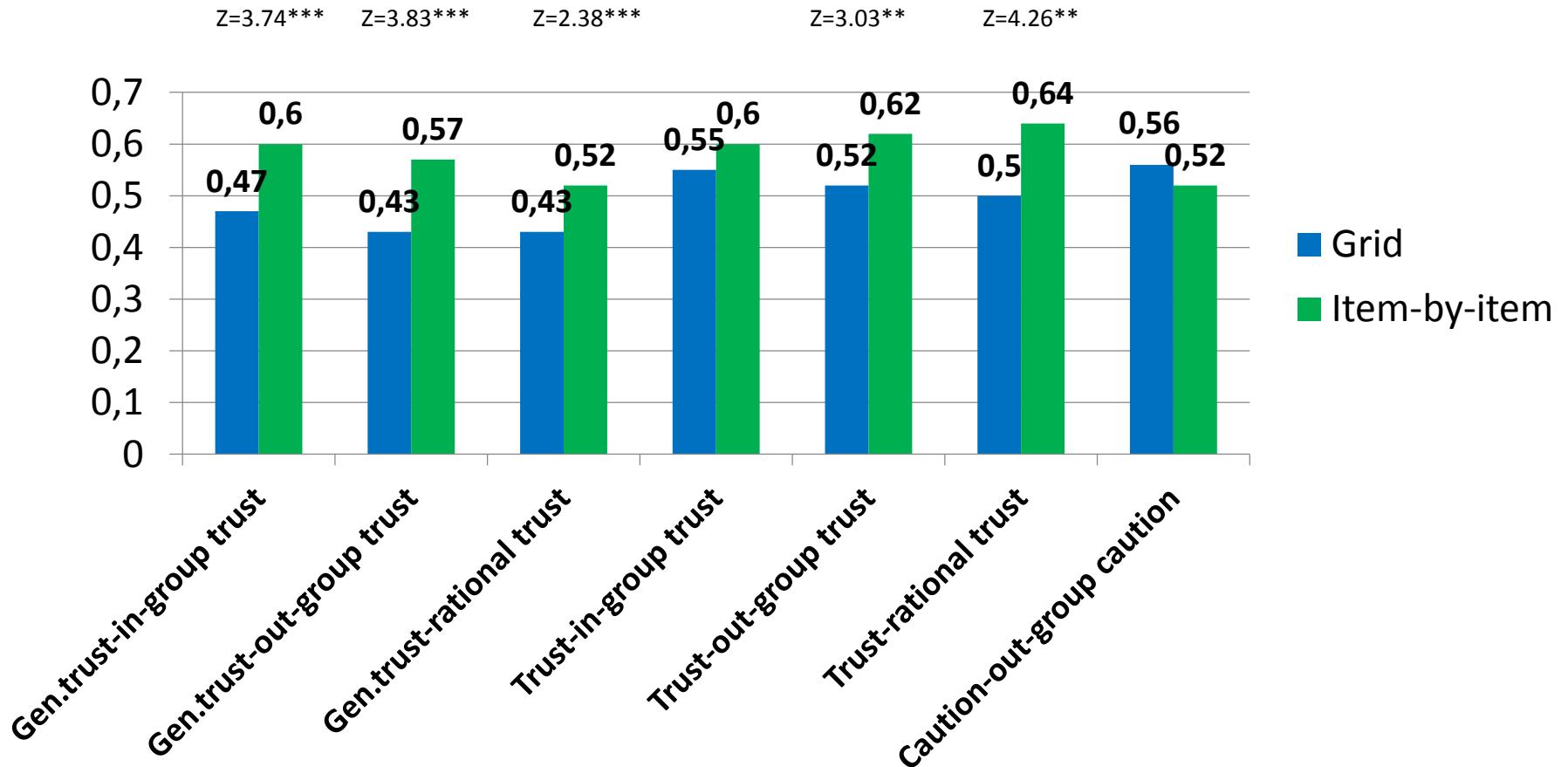
Grid format has lower concurrent validity

Wave 1, wave 2: similar but less evident pattern Wave 2

*** p<0.001, ** p<0.01, * p<0.05

3. Results: Concurrent validity (Device)

Z=3.74***   Z=2.14*

	PC	Mobile
Gen.trust-in-group trust	0,47	0,6
Gen.trust-out-group trust	0,47	0,53
Gen.trust-rational trust	0,46	0,49
Trust-in-group trust	0,54	0,61
Trust-out-group trust	0,55	0,59
Trust-rational trust	0,57	0,57
Caution-out-group caution	0,48	0,5

*** p<0.001, * p<0.05
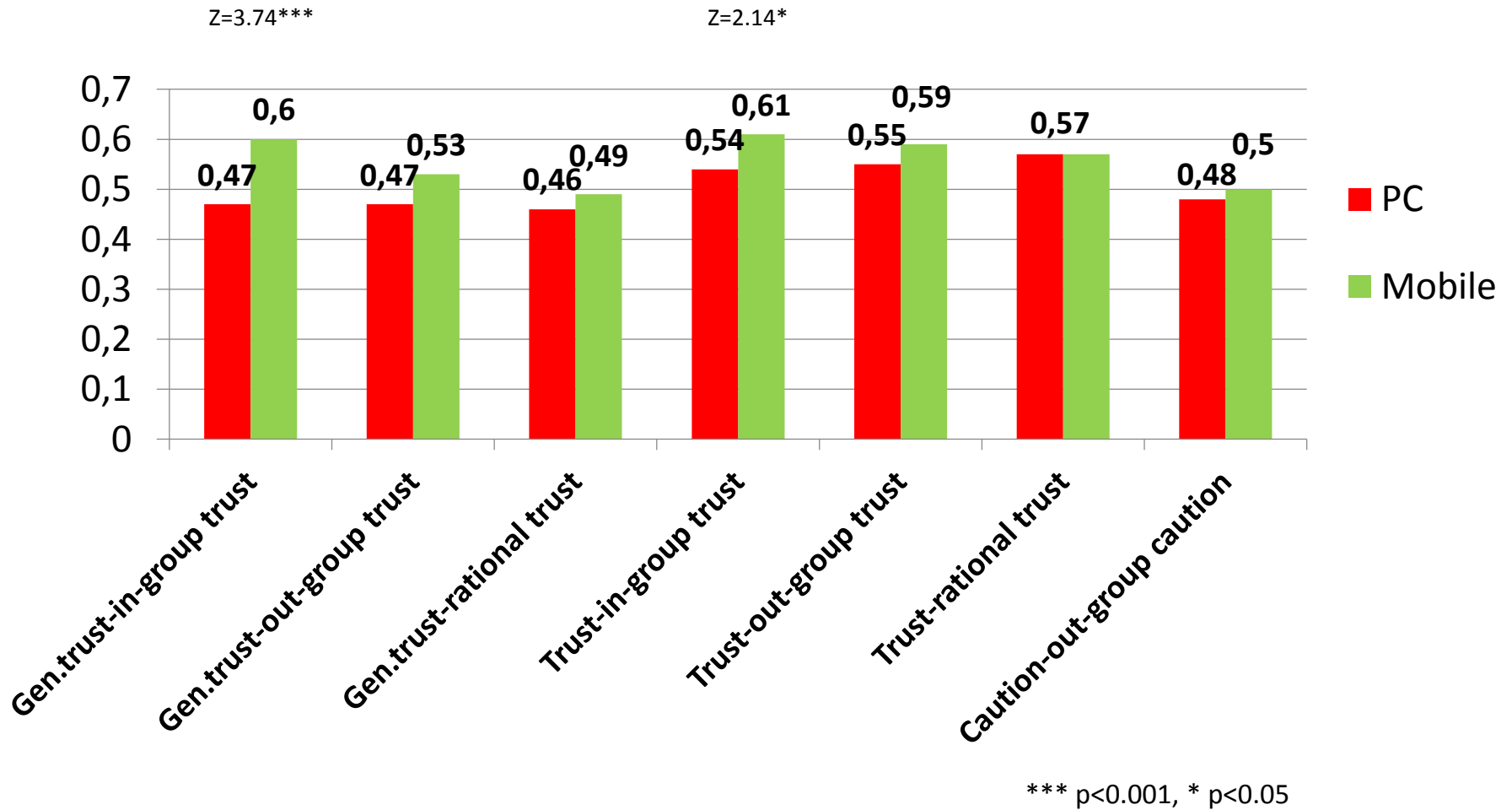
Wave 1

4. Results: Straightlining

	Straightlining (negative binomial model)	
	Wave 1	Wave 2
	Odds ratios	Odds ratios
Intercept	0.22*** [0.16-0.31]	0.25*** [0.16-0.38]
Grids (vs. item-by-item format)	1.34** [1.11-1.63]	1.38* [1.09-1.73]
PC (vs. mobile)	1.14 [0.94-1.38]	0.87 [0.69-1.09]
Males	1.32** [1.10-1.59]	1.17 [0.93-1.46]
Age	1.01 [1.00-1.02]	1.01 [1.00-1.02]
Model	Likelihood ratio $\chi2(4)=30.83$, p<0.001	Likelihood ratio $\chi2(4)=14.19$, p<0.01

No significant format*device interaction.

Grid format produced higher measurement error

5. Results: Measurement equivalence

	Number of latent factors	Response scale	Format (grid vs. item-by-item)	Device (PC vs. mobile)	Format X Device
Set 1: Risk willingness	1 factor	4-point scale	+	+	+
Set 2: Trust	2 factors	4-point scale	+	+	+
Set 3: Caution	2 factors	4-point scale	+	+	Configural (wave 1) Covariances of latent variables (wave 2)
Set 4: Moral and rational trust	2 factors	7-point scale	Residuals	+	Residuals (wave 1) Latent means (wave 2)
Set 5: Institutional trust	2 factors	5-point scale	+	+	+
Set 6: Tolerance	3 factors	4-point scale	+	+	+
Set 7: Schwartz values	2 factors	6-point scale	+	+	+

The measurement equivalence from the weakest to the strongest forms of measurement equivalence: 1) configural equivalence, 2) metric equivalence, 3) scalar equivalence, 4) latent means, 5) residuals, 6) variances of latent variables, 7) covariances of latent variables (when there are at least two latent factors).

Cells in red: the stronger form of measurement equivalence was not reached.
+: the stronger form of measurement equivalence was reached.

6. Results: Test-retest reliability

- We predicted the score of the latent variable in Wave 2 based on:
  - ❖ score of the latent variable in Wave 1,
  - ❖ question format: grid format in both waves, item-by-item format in both waves, and condition with different question formats between the waves
  - ❖ device
  - ❖ controlling for age and gender.
- Almost no differences between the question formats.
- No clear pattern of the differences between devices.
- No significant format*device interaction.

There is no clear answer which question format or device produced lower measurement error in terms of reliability

7. Results: Subjective indicators of respondent burden (odds ratios)
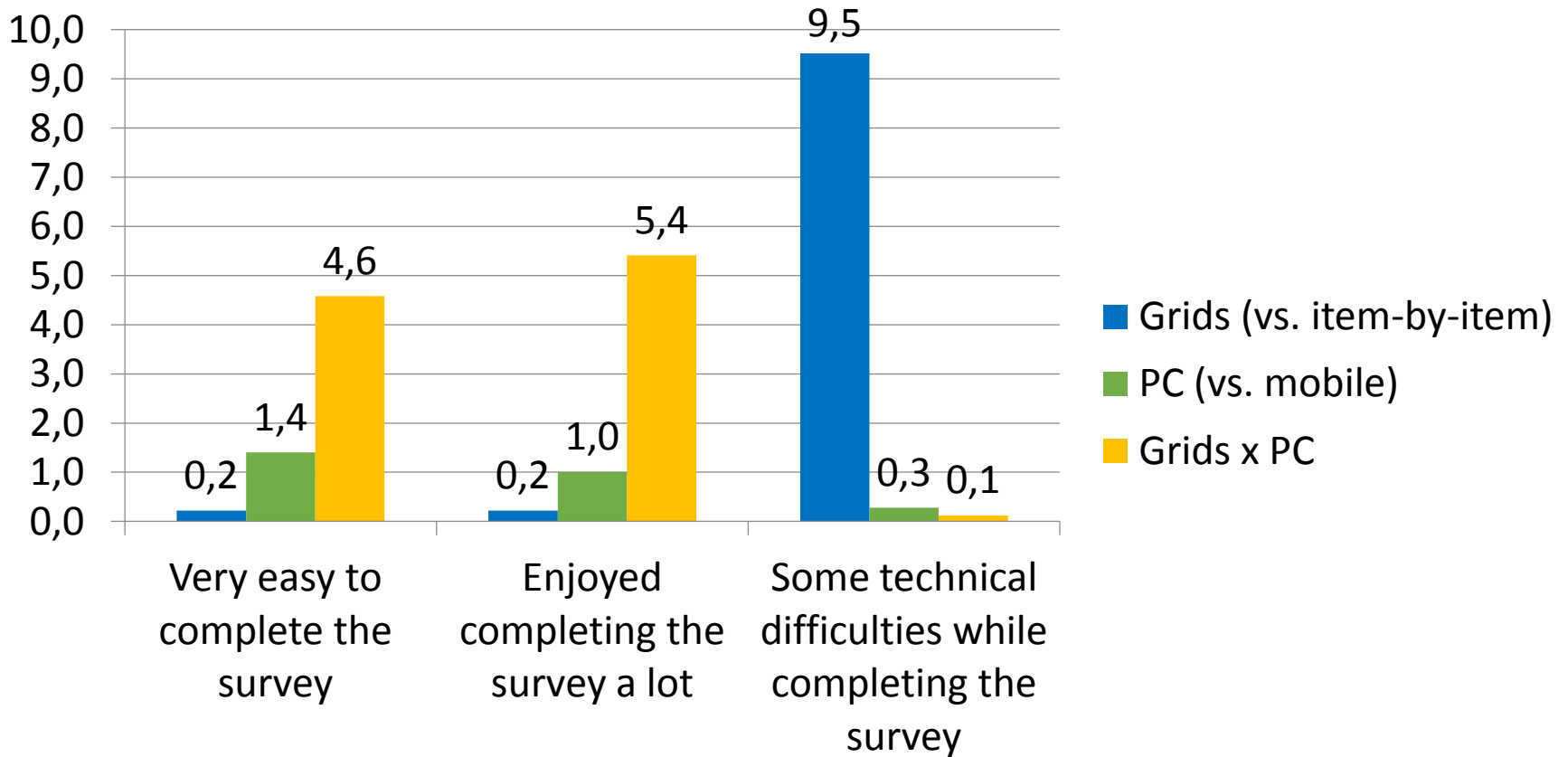
Wave 1, similar results in Wave 2

Main Findings

- ✓ Measurement error in the grid format was higher than in the item-by-item format on both PCs and smartphones:

  - lower concurrent validity

  - higher straightlining

- No significant effects of the question format on test-retest correlations.

- No differences in breakoff rates between the question formats. Mobile web produced a higher breakoff rate, but no differences between the question formats across devices.

- The grid format had shorter completion times on both devices; however, at the cost of higher measurement error. Subjective evaluation of interview length was longer in the grid format. The effect was strong among smartphone respondents.

- Subjective indicators of respondent burden: the grid format decreased subjective evaluation of the survey and increased reported technical difficulties. The effect was substantial in the mobile web condition (but the condition was non-optimized).

Practical Implications

- If one is not able to use mobile optimization for smartphones, using an item-by-item format for mobile web surveys may be prudent. Using grids on mobile devices without optimization increases measurement error and decreases the subjective evaluation of the survey.

- Using the item-by-item format in PC web surveys may increase concurrent validity and decrease non-differentiation or straightlining. Although the grid format decreased survey completion time, it had no effect on subjective evaluation of the survey length among PC web respondents.

- In questions with 7 or more response options we recommend using an item-by-item format on both devices; otherwise, there may be differences in measurement equivalence between devices.

The choice of whether or not to use a grid format should be made on a case-by-case basis, and is not an all-or-nothing decision.

Thank you for your attention!

Aigul Mavletova

amavletova@hse.ru