Comparing the Performance of AD and IS Questions across PCs and Smartphones

JAN KAREM HÖHNE[1], MELANIE REVILLA[2], TIMO LENZNER[3]

[1]UNIVERSITY OF GÖTTINGEN

[2]RECSM-UNIVERSITAT POMPEU FABRA

[3]GESIS – LEIBNIZ INSTITUTE FOR THE SOCIAL SCIENCES

72nd AAPOR Conference

18 - 21 May 2017 – New Orleans, LA (USA)

Survey Practice

- Since Likert (1932) Agree/Disagree (AD) questions have become popular.[1]
  - *German General Social Survey, Eurobarometer, ISSP …*
  - *European Sociological Review, Political Analysis, Social Science Quarterly …*
- Reasons for popularity[2]:
  - Measuring different contents with the same scale.
  - Streamlining questionnaires – using grids.
  - Reducing administration time.
  - Simpler to design.

[1] Revilla (2016)
[2] Saris et al. (2010)

Critique on Survey Practice

- Fowler (1995) identified several drawbacks of AD questions:
  - Insufficient anchoring.
  - Complex cognitive processing.
  - Small discriminatory power.
  - Prone to response bias (e.g., acquiescence).

Item-Specific (IS) questions represent a simpler, more informative, and more direct method (Fowler 1995).

Research Question

- Do AD questions require more cognitive effort than IS questions, regardless of the device?
  - So far, empirical evidence is mixed.

What are possible REASONS for differences?

Concept of "Asking Manner"

- AD questions do not change *"asking manner"*.
  - Fostering a superficial response process.
  - Indirect statement.
- IS questions change *"asking manner"* permanently.
  - Require constant reconsideration of the dimension of interest.
  - Encourage an active and intensive response process.
  - Direct question.

Accordingly, IS questions seem to be more demanding than AD questions.

Höhne et al. (in press)

Research Hypotheses

IS questions produce longer response times than AD questions. (H1)

The response time differences are higher in PCs than smartphone. (H1a)

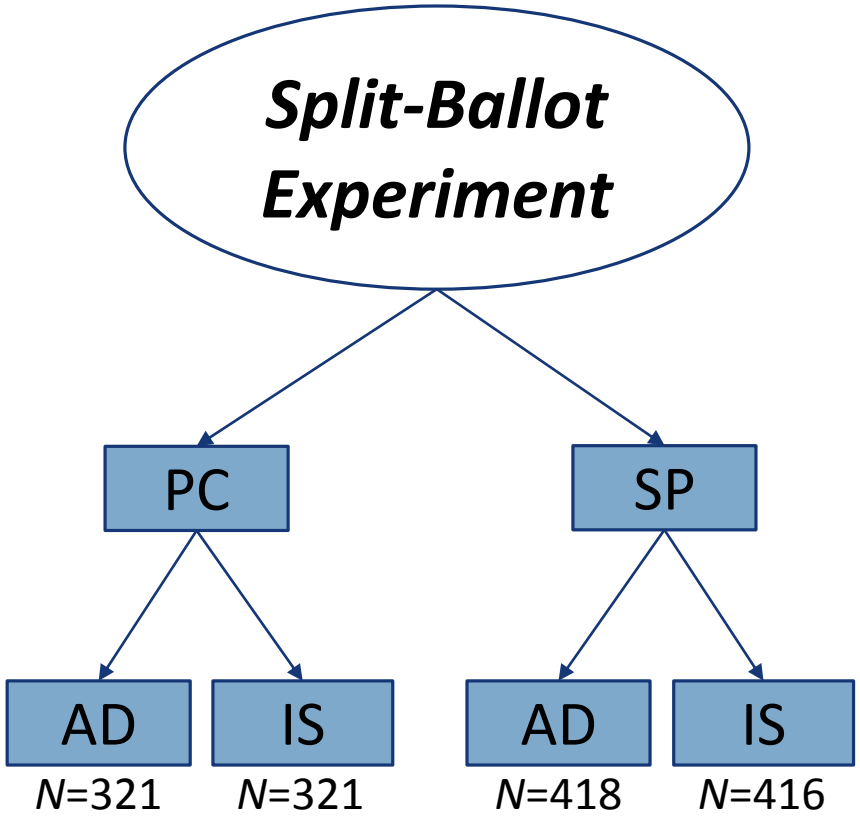
The response time differences are higher in 7- than 5-point scales. (H1b)

IS questions produce higher response quality than AD questions. (H2)

The response quality differences are higher in PCs than smartphone. (H2a)

The response quality differences are higher in 7- than 5-point scales. (H2b)

Research Design



- Population of interest: Netquest panelists with Internet access through PC and smartphone.
- Participants were first randomly assigned to a device
  - Based on profiling information.
  - Detection of wrong device.
- Then, they were randomly assigned to a question format.
- Respondents could skip questions.

Survey Questions

AD Question:

I am relaxed most of the time.

agree strongly

…

disagree strongly

IS Question:

How often are you relaxed?

very often

…

never

All groups received six 5- and then six 7-point scale questions dealing with personality traits. Scales were end-labeled and vertically arranged.

Sample

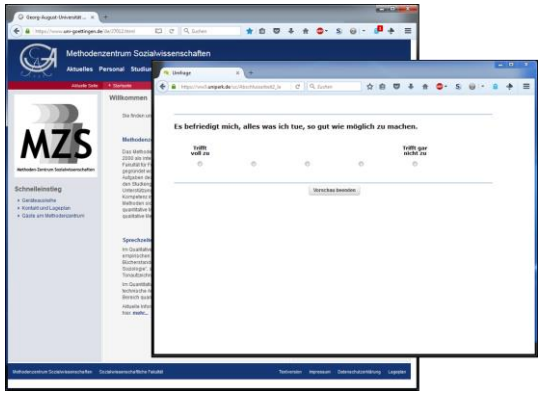
The web survey was conducted by the opt-in access panel Netquest in Spain from September to October 2016.

Cross quotas:	Gender and Age
Final sample size:	N = 1,476
Gender:	61% female
Age:	Mean = 36.4 (SD = 11.8)
Education:	1% lower secondary school 26% intermediate secondary school 73% at least college preparatory secondary school
Spanish native:	93% native speakers

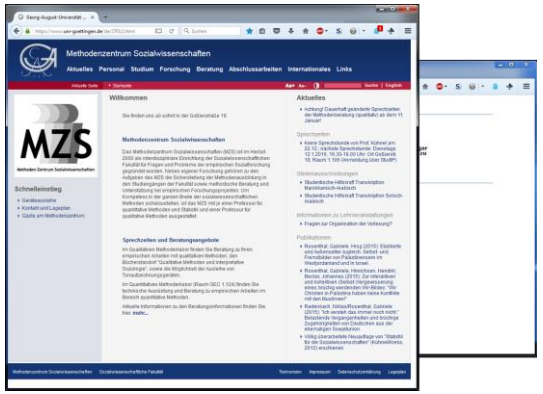
Outlier Definition Procedure

1st Step

Participants who left the web-survey page were defined as outliers.



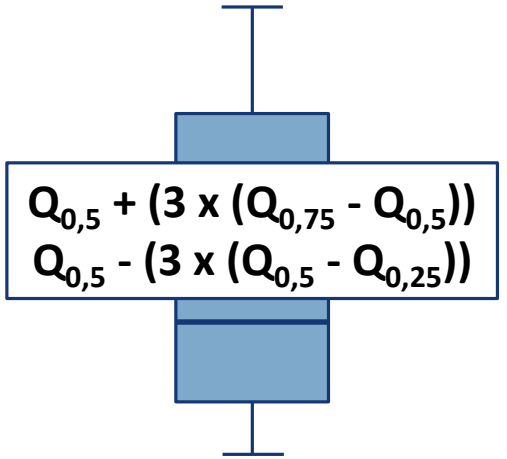
SurveyFocus: ON



SurveyFocus: OFF

2nd Step

"Common" response time outlier definition for remaining participants.



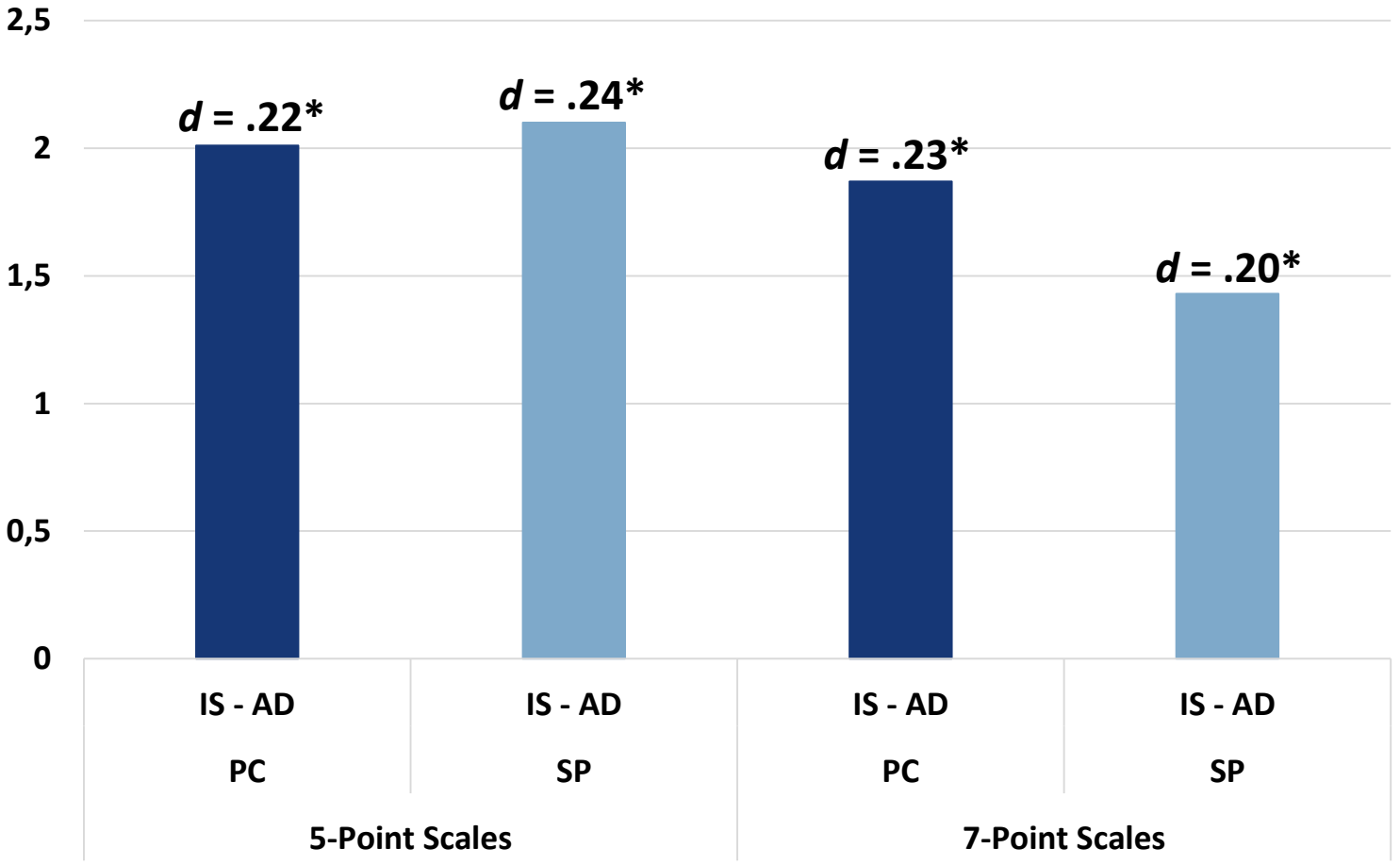
Höhne & Schlosser (2017)
Höhne et al. (in press)

Analytical Strategy

- Response Times:
  - No checking for baseline reading speed.[1]
  - Response times divided by syllables.[2]
- Response Quality:
  - Speeding → extremely fast responding.
  - Primacy effects → attraction to first category.
  - We checked further indicators but found no differences.
- Analytical level:
  - Aggregation of the 5- and 7-point scale questions.
- Robustness checks: All results remained unchanged.

[1] Couper & Kreuter (2013)
[2] Lenzner et al. (2010)

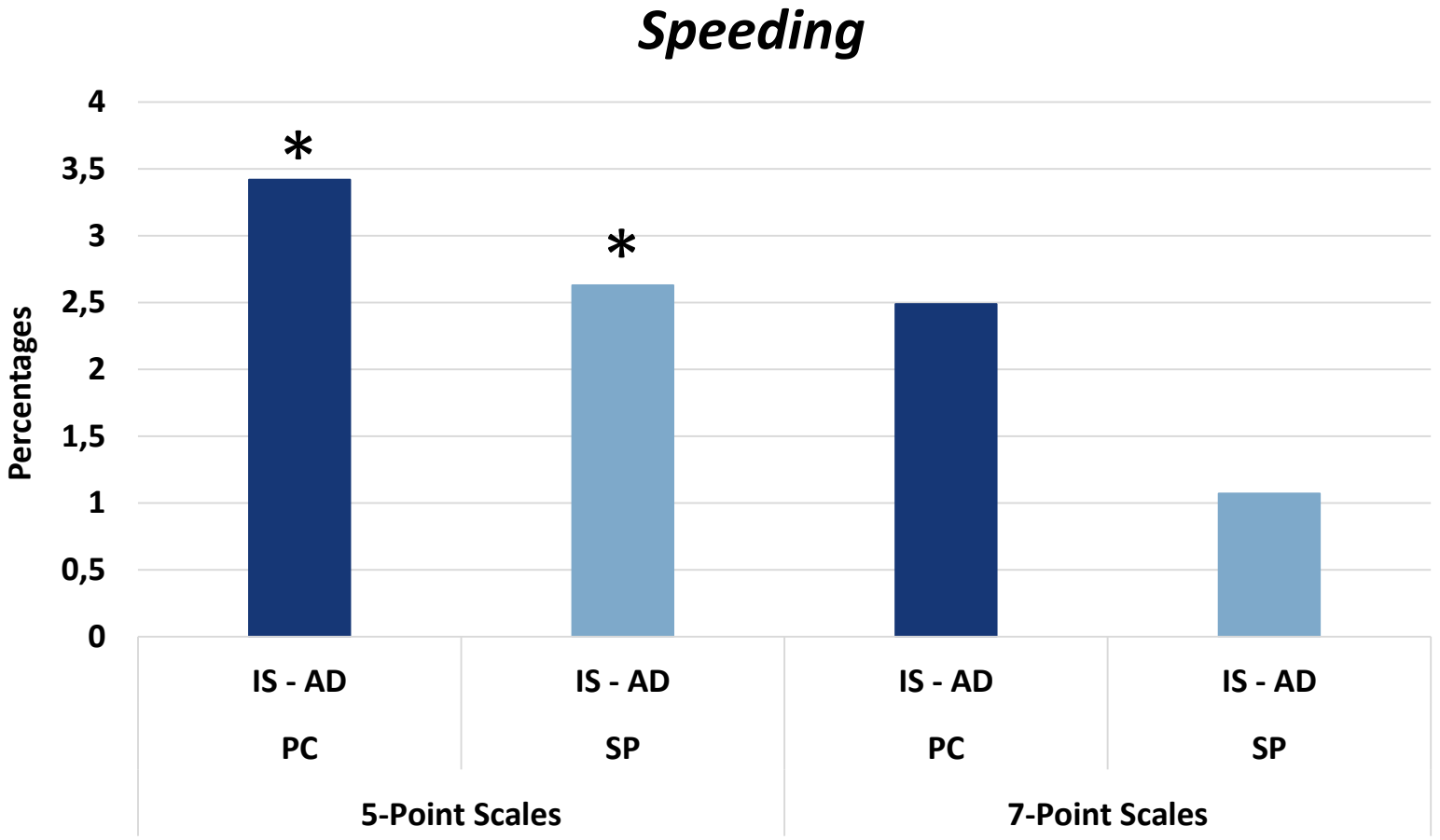
Results: Response Time Differences



Hypotheses H1a & H1b
5-Point (PC vs. SP): $z > -1.96$
7-Point (PC vs. SP): $z < 1.96$
PC (5- vs. 7-Point): $z < 1.96$
SP (5- vs. 7-Point): $z < 1.96$

***Note.** *p < .05. F-test. Response time differences per syllable are reported in the graph. Mean differences: IS group minus AD group. Cohen's d indicates the effect size.*

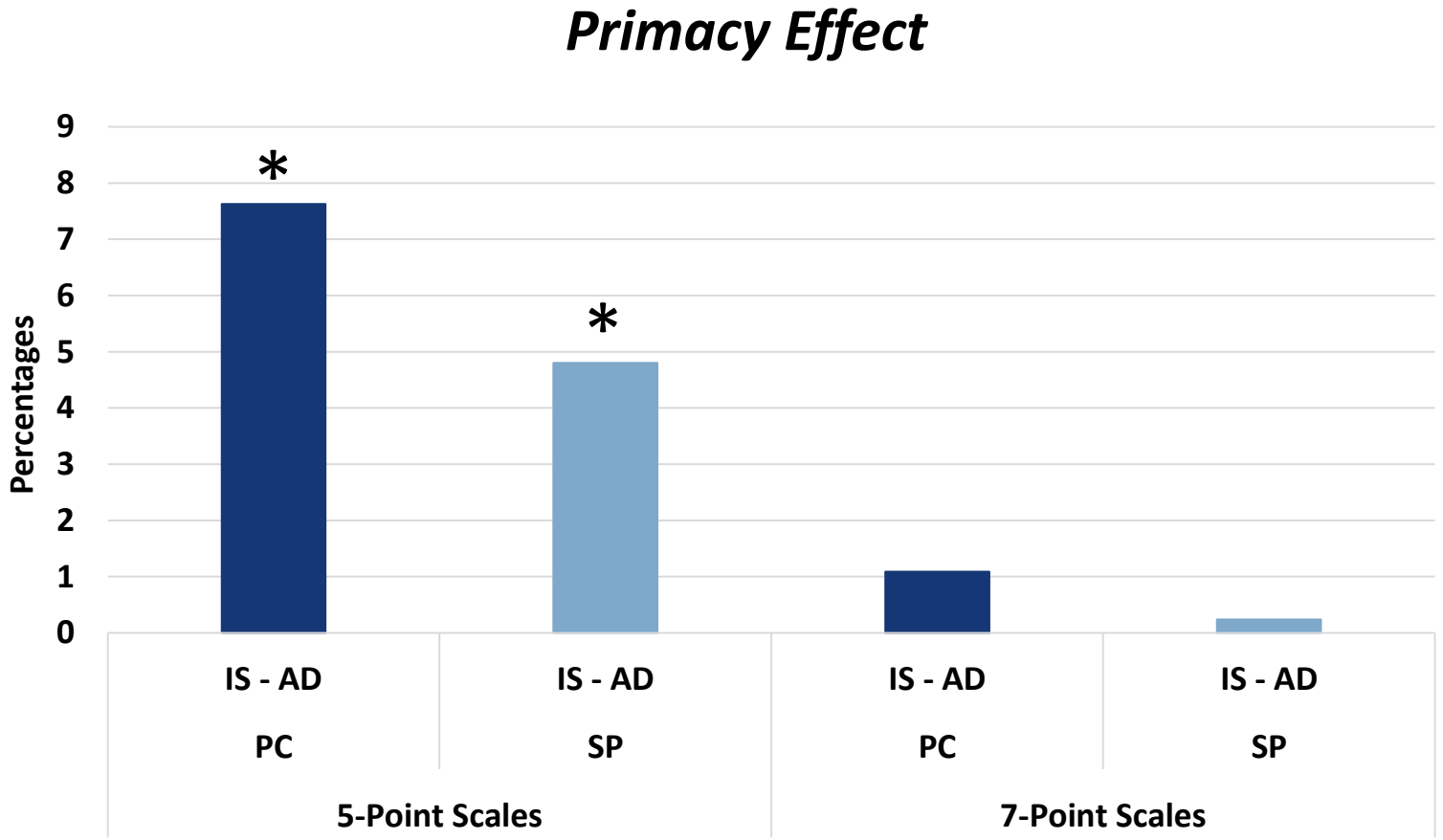
Results: Response Quality Differences I



Hypotheses H2a & H2b
5-Point (PC vs. SP): $z < 1.96$
7-Point (PC vs. SP): $z < 1.96$
PC (5- vs. 7-Point): $z < 1.96$
SP (5- vs. 7-Point): $z < 1.96$

*Note. *p < .05. Chi-square tests. For speeding we used the lower 15th percentile of all response times. Differences: AD group minus IS group.*

Results: Response Quality Differences II



Hypotheses H2a & H2b
5-Point (PC vs. SP): $z < 1.96$
7-Point (PC vs. SP): $z < 1.96$
PC (5- vs. 7-Point): $z > 1.96$*
SP (5- vs. 7-Point): $z > 1.96$*

*Note. *p < .05. Chi-square tests. For primacy effects we used the number of responses given to the first category. Differences: AD group minus IS group.*

Reflection

- Limitations
  - Specific target population.
  - Trained respondents.
  - Only one country.

- Future research
  - Checking robustness: Population, survey experience, country.
  - Question topics, arrangement/direction of the scale, and scale labeling.

Summary & Conclusion

- Question Format influence response times. (H1) ✓
  - IS questions seem to be more demanding → higher response times.
- IS questions produce higher response quality. (H2) ✓
  - Speeding and Primacy effects → 5-point scales.
- No clear pattern regarding differences between devices and scale lengths. (Ha & Hb, resp.) ×
- Conclusion: Differences between presumed cognitive complexity and expended cognitive effort.

Finally, we recommend to employ IS instead of A/D questions.

Thank you for your attention!

Contact: jhoehne@uni-goettingen.de

Literature

- Couper, M. P., & Kreuter F. (2013). Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society*, 176, 271–286.
- Fowler, F. (1995). Improving Survey Questions: Design and Evaluation. Thousand Oaks, CA: Sage.
- Höhne, J. K., & Schlosser, S. (2017). Investigating the Adequacy of Response Time Outlier Definitions in Computer-based Web Surveys using Paradata SurveyFocus. *Social Science Computer Review*.
- Höhne, J. K., Schlosser, S., & Krebs, D. (in press). Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys using Paradata. *Field Methods*.
- Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment. *Applied Cognitive Psychology*, 24, 1003–1020.
- Revilla, M. (2016). Analyzing the Survey Characteristics, Participation, and Evaluation across 186 Surveys in an Online opt-in Panel in Spain. *Paper presented at the QDET2 conference*, Miami, FL (USA).
- Saris, W., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. *Survey Research Method*, 4, 61-79.