Surveys or digital trace data, which one should we use?

Using MultiTrait-MultiMethod models to simultaneously estimate the measurement quality of surveys and digital trace data.

Oriol J. Bosch| University of Oxford, LSE & RECSM

Melanie Revilla |Institut Estudis Barcelona Internacional (IBEI)

Patrick Sturgis| Department of Methodology, LSE

Jouni Kuha| Department of Methodology, LSE

o.bosch-jover@lse.ac.uk

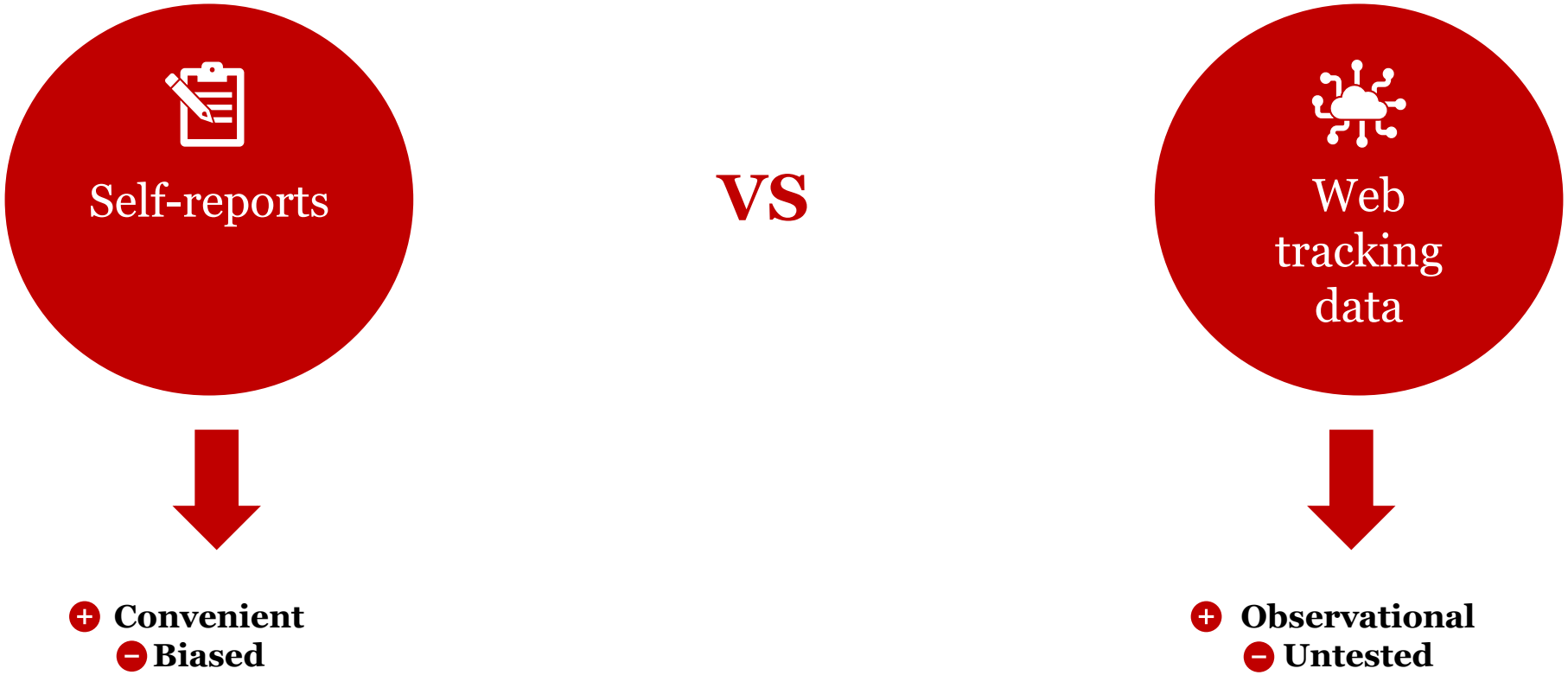
orioljbosch

https://orioljbosch.com/

Funding: This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 849165; PI: Melanie Revilla); the Spanish Ministry of Science and Innovation under the "R+D+i projects" programme (grant number PID2019-106867RB-I00 /AEI/10.13039/501100011033 (2020-2024), PI: Mariano Torcal); and the BBVA foundation under their grant scheme to scientific research teams in economy and digital society, 2019 (PI: Mariano Torcal).

Survey or digital trace data, that is the question

The emergence of digital trace data means that now we can directly observe what people do online



Web tracking data

Direct observations of online behaviours using tracking solutions, or *meters.*

Group of tracking technologies (plug-ins, apps, proxies, etc)

Installed on participants devices

Collect traces left by participants when **interacting with their devices online: URLs, apps visited, cookies…**

Great, we will get unbiased measures!

Web tracking is not perfect!

Web tracking data, as any other data source, **is affected by errors**

JOURNAL ARTICLE

When Survey Science Met Web Tracking: Presenting an Error Framework for Metered Data

Oriol J. Bosch ✉, Melanie Revilla Author Notes

Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 185, Issue Supplement_2, December 2022, Pages S408–S436, https://doi.org/10.1111/rssa.12956

Published: 06 November 2022 **Article history** ▾

Validity and Reliability of Digital Trace Data in Media Exposure Measures: A Multiverse of Measurements Analysis

Oriol J. Bosch[1, 2, 3]

[1] Leverhulme Centre for Demographic Science, University of Oxford
[2] Department of Methodology, The London School of Economics and Political Science
[3] Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra

Uncovering digital trace data biases: tracking undercoverage in web tracking data

Oriol J. Bosch[1, 2, 3], Patrick Sturgis[2], Jouni Kuha[2], Melanie Revilla[4]

[1] Leverhulme Centre for Demographic Science, University of Oxford
[2] Department of Methodology, The London School of Economics and Political Science
[3] Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra
[4] Institut Barcelona Estudis Internacionals (IBEI)

Web tracking is not perfect!

Web tracking data, as any other data source, **is affected by errors**

JOURNAL ARTICLE

When Survey Science Met W
an Error Framework for Me

Oriol J. Bosch ✉, Melanie Revilla Author Not

Journal of the Royal Statistical Society Series A:
Supplement_2, December 2022, Pages S408–S4
Published: 06 November 2022 **Article hist**

VALIDITY AND RELIABILITY OF DIGITAL TRACE DATA IN MEDIA
OF MEASUREMENTS ANAL-

3
ence, University of Oxford
of Economics and Political Science
dology, Universitat Pompeu Fabra

To make a proper informed decision, we should compare the actual quality of both data sources when measuring the same concepts

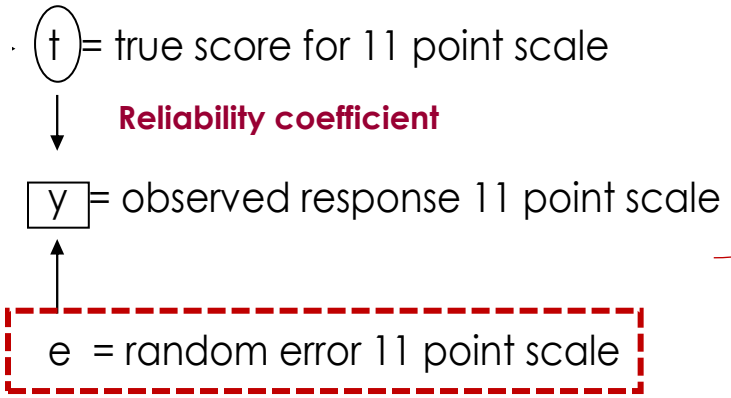
UNCOVERING DIGITAL TRACE DATA BIASES: TRACKING UNDERCOV-
ERAGE IN WEB TRACKING DATA

Oriol J. Bosch[1, 2, 3], Patrick Sturgis[2], Jouni Kuha[2], Melanie Revilla[4]

[1] Leverhulme Centre for Demographic Science, University of Oxford
[2] Department of Methodology, The London School of Economics and Political Science
[3] Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra
[4] Institut Barcelona Estudis Internacionals (IBEI)

Simultaneously estimating the measurement quality of digital trace data and surveys using MultiTrait-MultiMethod (MTMM) models

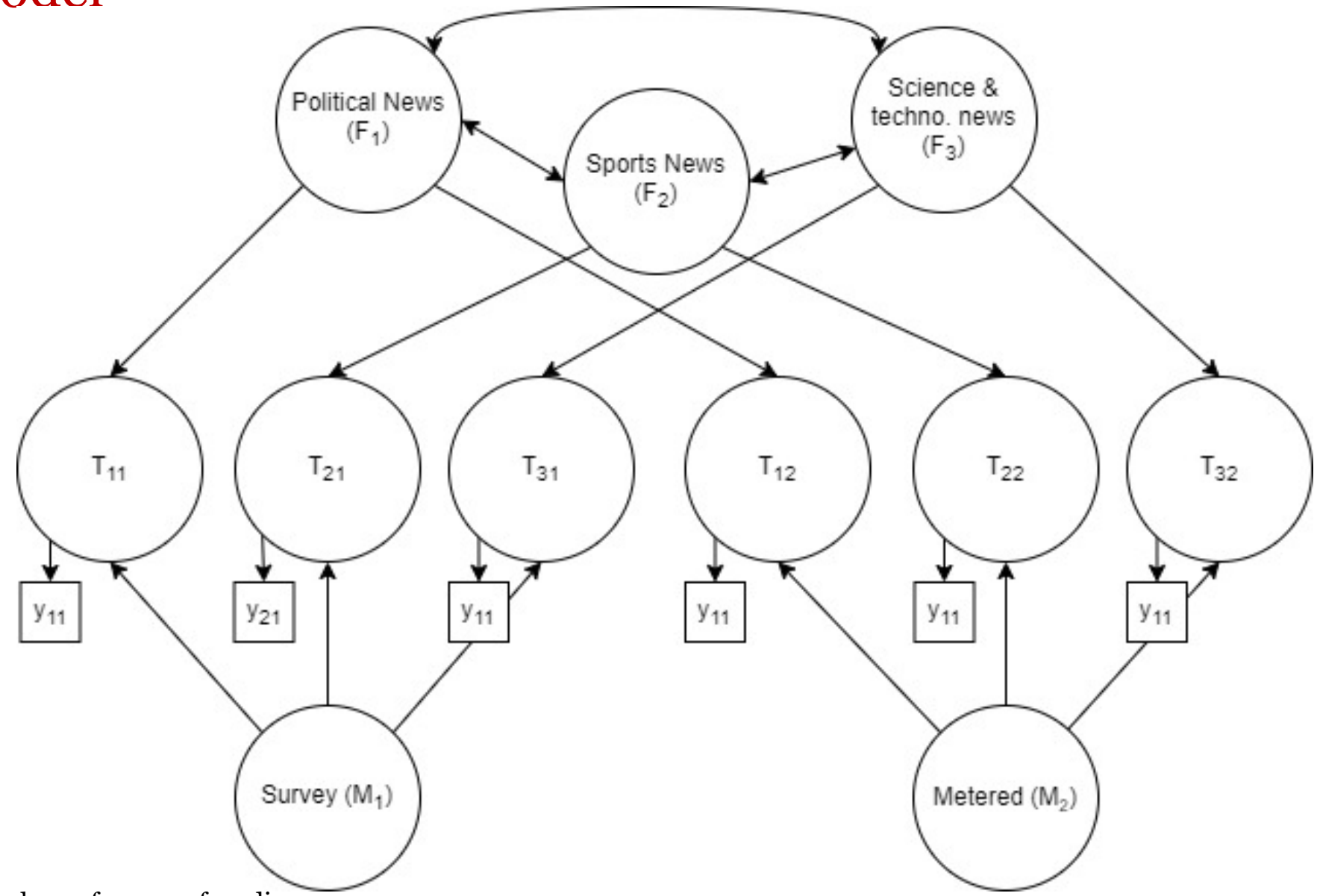
Measurement quality



Quality = reliability *x* measurement validity

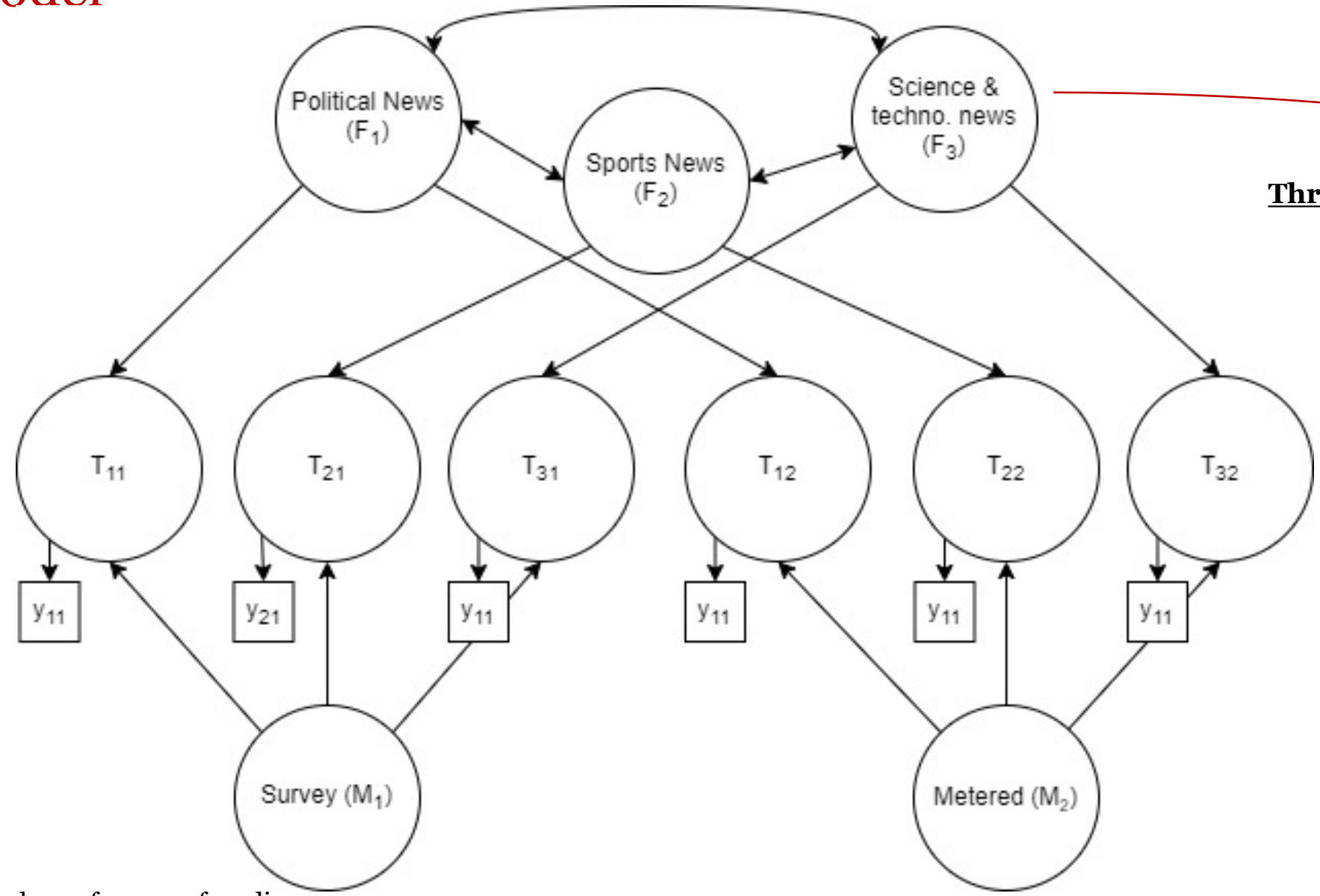
Quality = strength of the relationship between the latent concept of interest and the observed answers

The model



Residuals are not shown for ease of reading

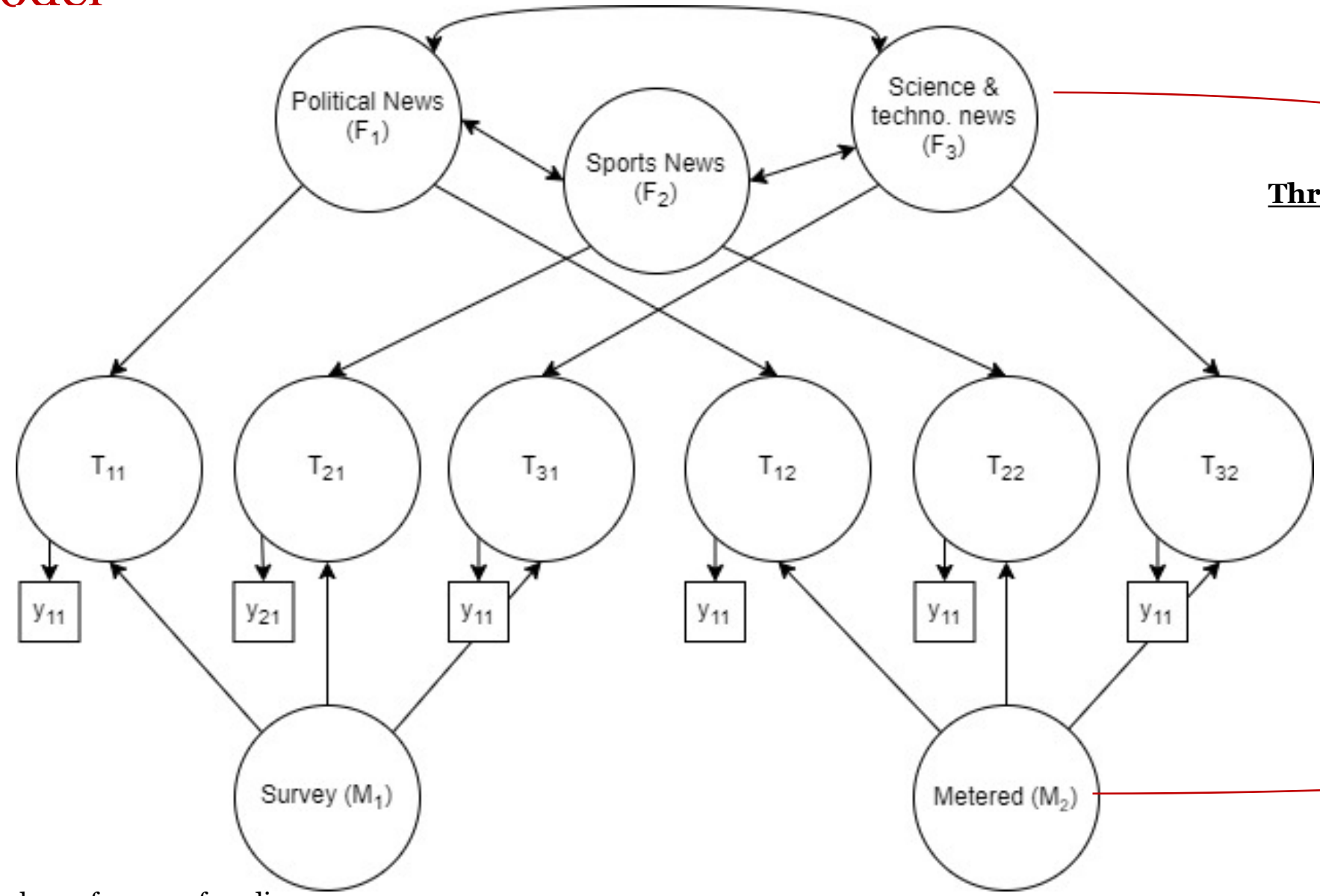
The model



Three similar traits

Residuals are not shown for ease of reading

The model

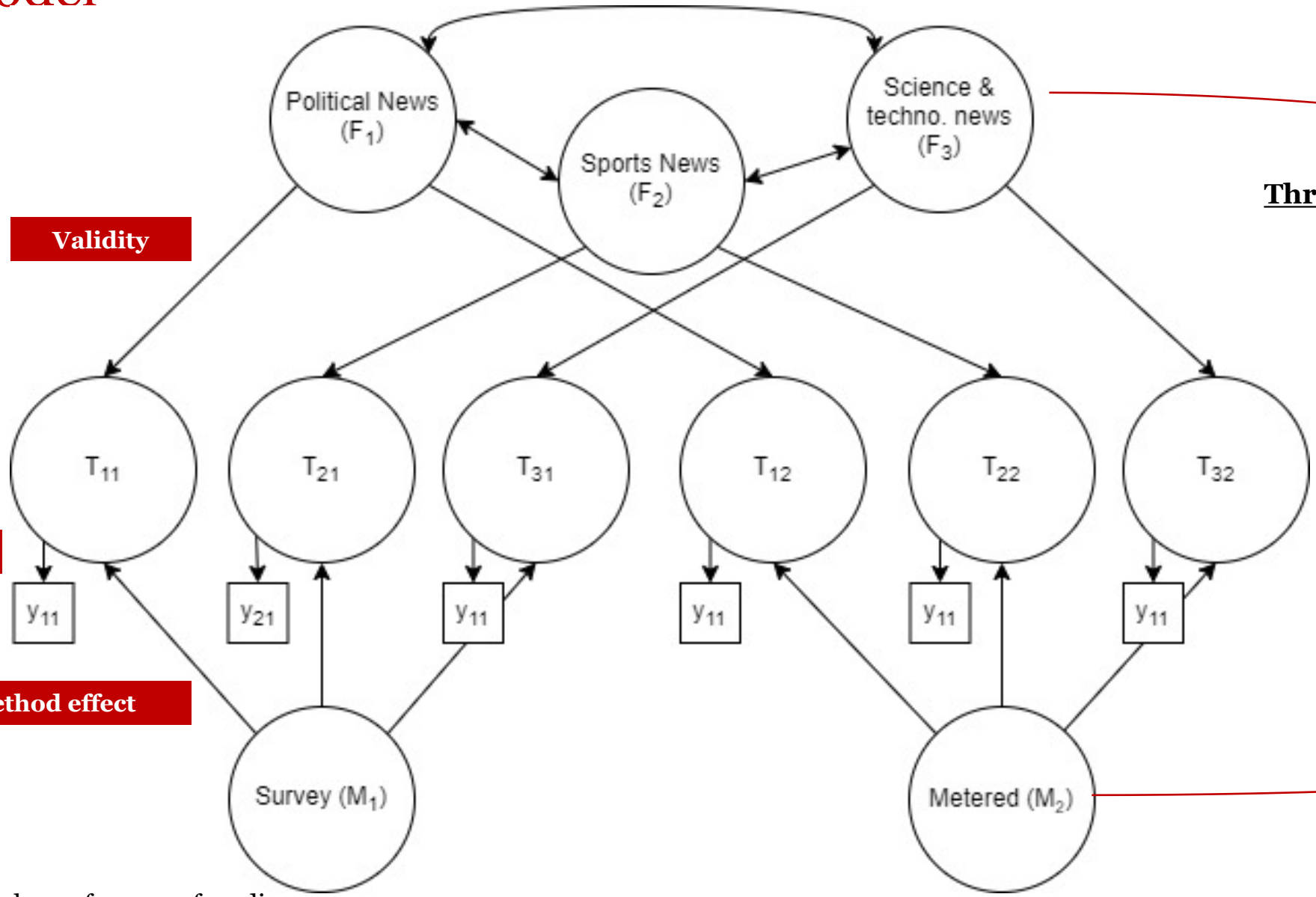


Three similar traits

Two methods to measure these traits

Residuals are not shown for ease of reading

The model



Validity

Reliability

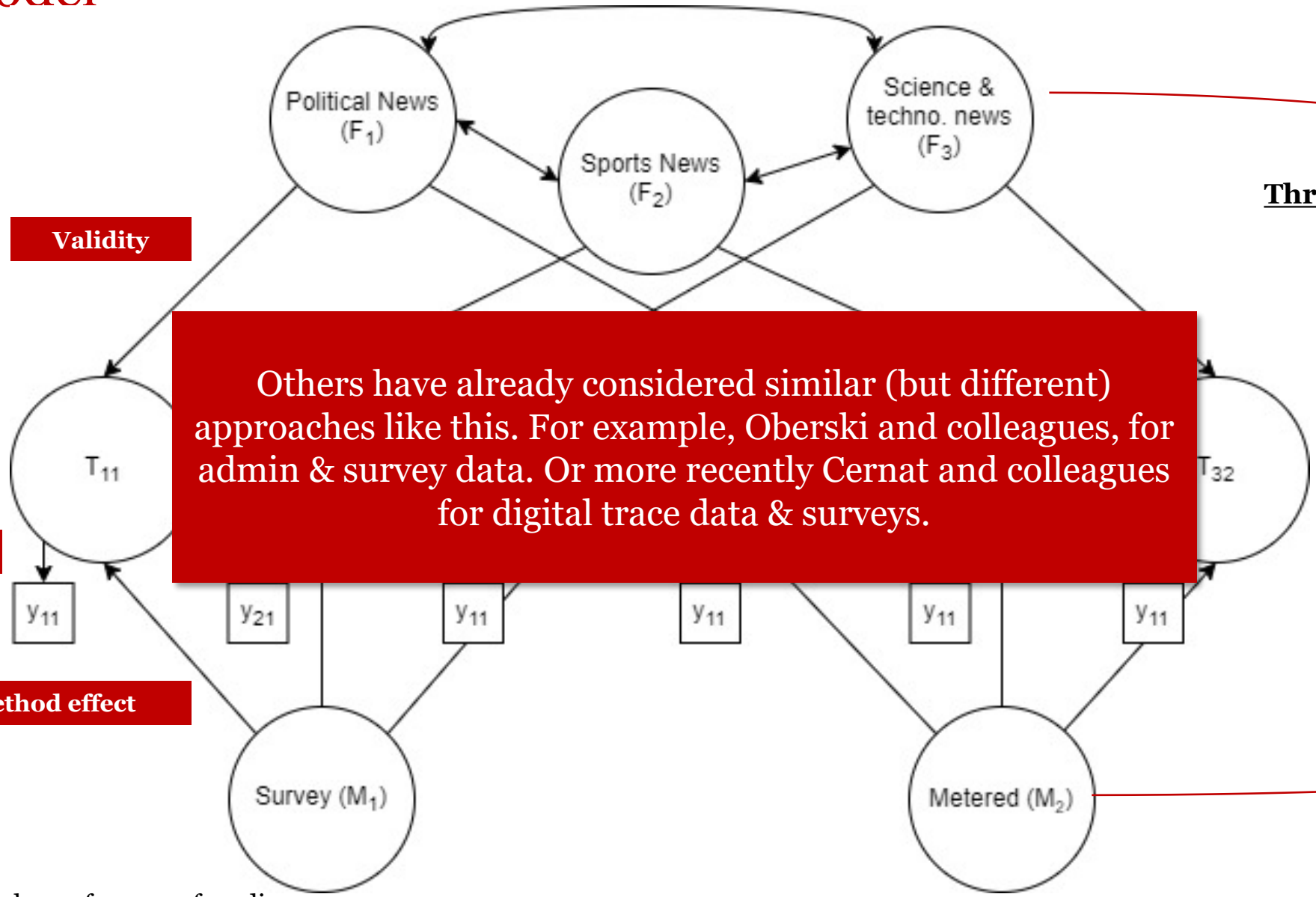
Method effect

Three similar traits

Two methods to measure these traits

Residuals are not shown for ease of reading

The model



Validity

Reliability

Method effect

Others have already considered similar (but different) approaches like this. For example, Oberski and colleagues, for admin & survey data. Or more recently Cernat and colleagues for digital trace data & surveys.

Three similar traits

Two methods to measure these traits

Residuals are not shown for ease of reading

This study

THIS STUDY

Data

- **Survey** combined with **web tracking data** at the individual level

- Netquest metered panel in Spain
  - **Cross-quotas:** gender, age, and education
  - **Sample size:** 1,200
  - **Fieldwork:** Late May – Early June 2023

- Tracking technologies installed in both **mobile and desktop devices**

- Part of the ERC project **WEB DATA OPP**

THIS STUDY

Three differ groups of traits of interest

1. News exposure traits

- Exposure to news about politics
- Exposure to news about sports
- Exposure to news about science and technology

2. Communication traits:

- Use of social media
- Use of instant messaging
- Use of e-mails

3. Entertainment traits:

- Use of video platforms (YouTube, Vimeo, Twitch)
- Use of audio streaming (Spotify, Audible, Apple podcast)
- Use of TV/Movie streaming (Netflix, BBC online)

THIS STUDY

The measurements

1. Survey questions

More specifically, on average, how much time per day do you spend on the Internet reading news and articles...

- **MC4_1.** ... about politics and current affairs?

- **MC4_1_HH.** Hours: *[SMALL NUMERIC OPEN BOX]* **MC4_1_MM.** Minutes: *[SMALL NUMERIC OPEN]*

The measurements

1. Survey questions

2. Web tracking data

Characteristics	My choices
Metric	Minutes
List of traces	
List of media	Tranco
Top media	All
Information	Those identified as specific concept
Exposure	
Time threshold	1 second
Devices	All devices (with or without app)
Tracking period	31 days

THIS STUDY

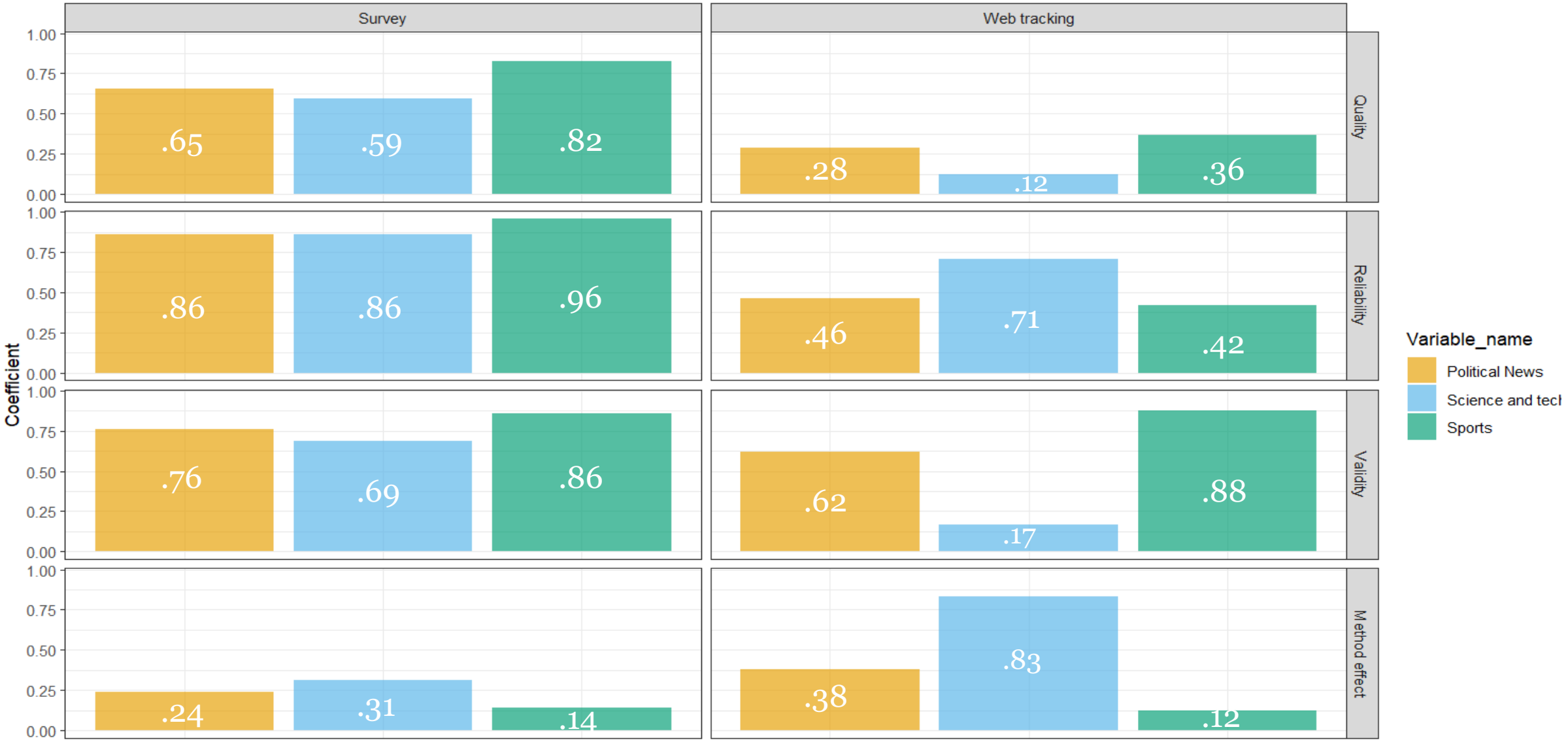
The measurements

1. **Survey questions**
2. **Web tracking data**

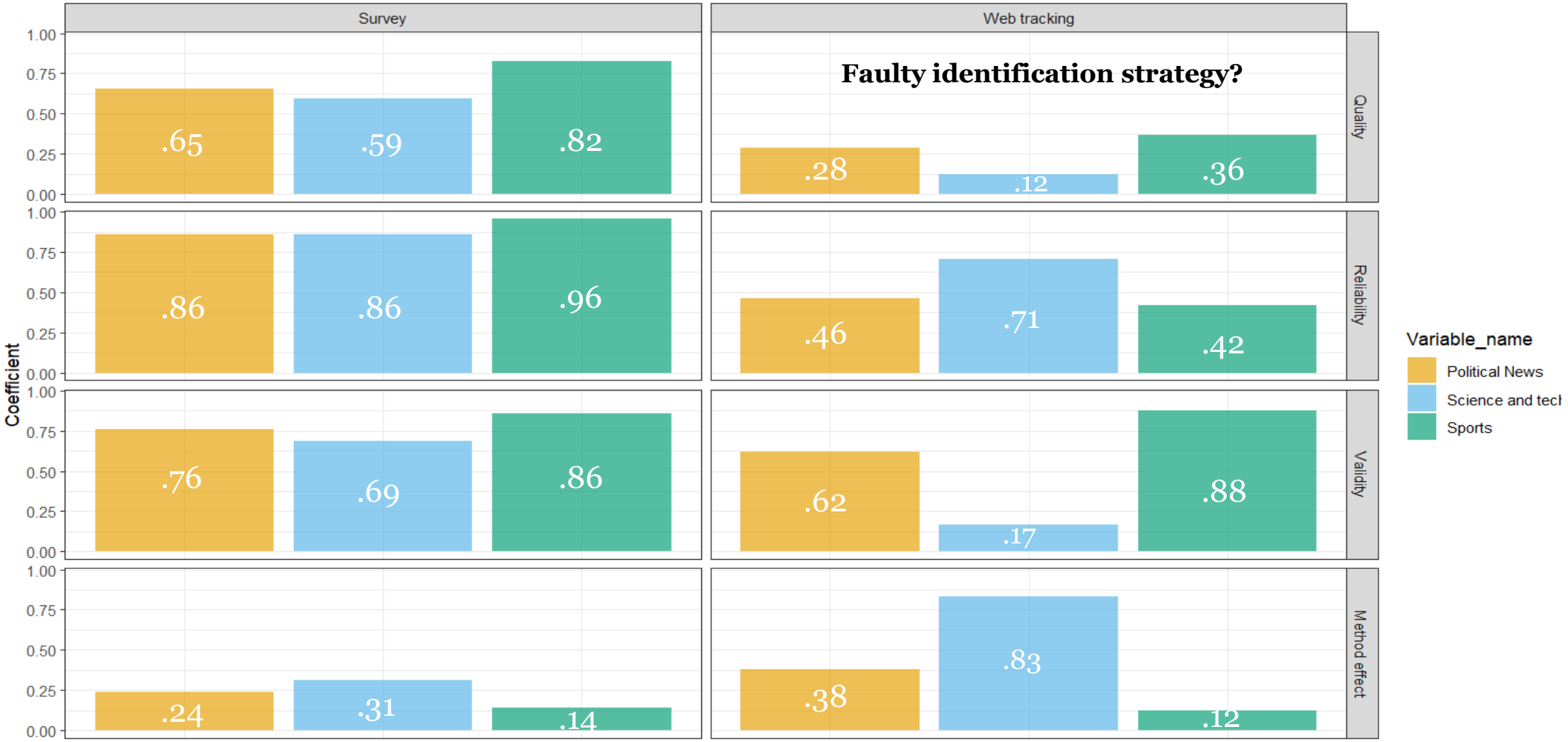
I use the log of these measures

Results

#1 News: quality estimates

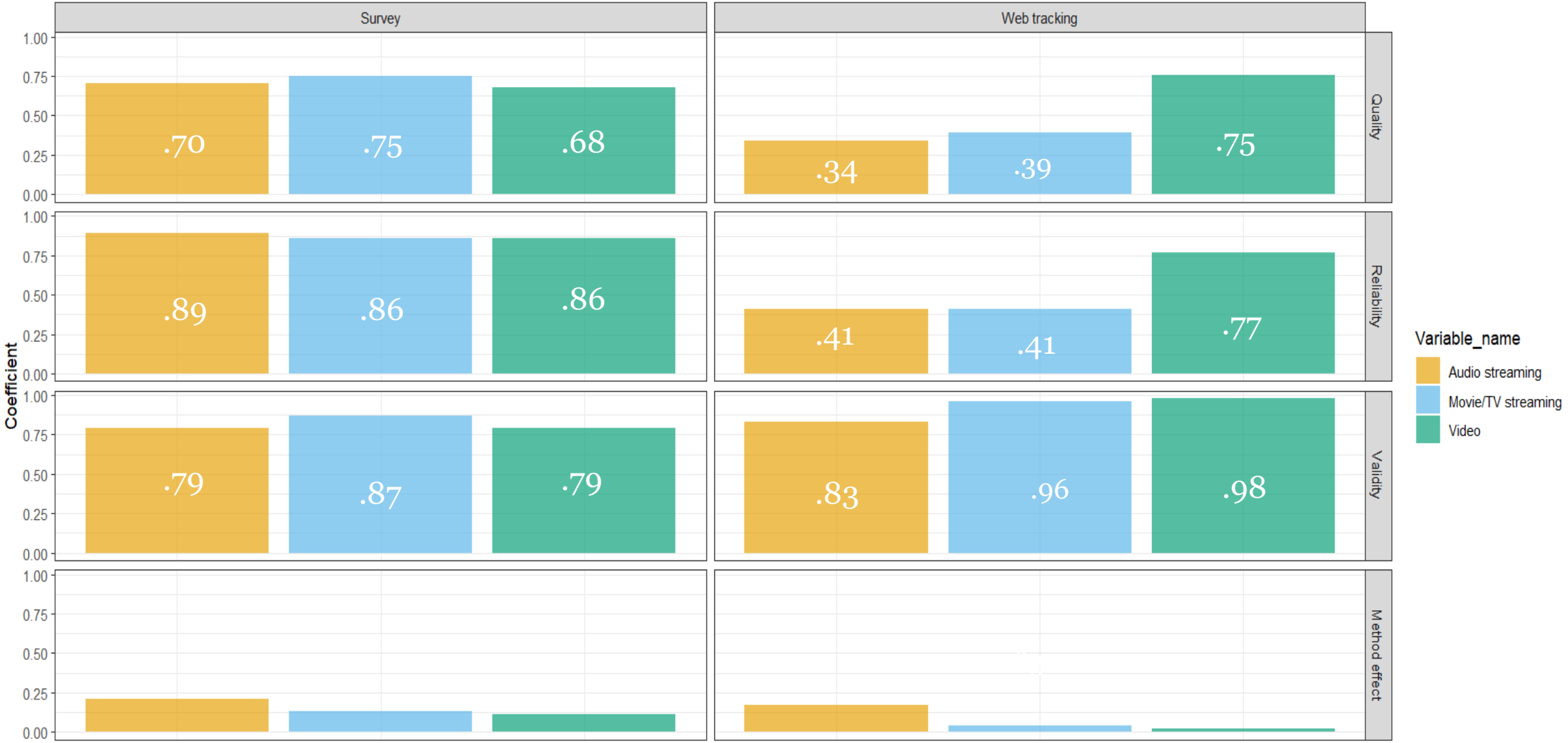


#1 News: quality estimates



#2 Communication: quality estimates

#3 Entertainment: quality estimates



RESULTS

#3 Entertainment: quality estimates

CONCLUSIONS

Take-home messages

- Results **put into question** the measurement quality of web tracking measurements
  - Some concepts are measures very accurately: **communication and video streaming**

    → **Variance explained by trait: +/- 80%**
  - While others are extremely off: **news media exposure and some entertainment**

    → **Variance explained by trait: 12-39% !!!**

Take-home messages

- Results **put into question** the measurement quality of web tracking measurements
  - Some concepts are measures very accurately: **communication and video streaming**

    ➡ **Variance explained by trait: +/- 80%**

  - While others are extremely off: **news media exposure and some entertainment**

    ➡ **Variance explained by trait: 12-39% !!!**

Surveys, on the other hand, perform acceptably well. They also struggle more with news, but their quality is never below .50 and generally around .70 (agrees w/ Alwin)

CONCLUSIONS

Take-home messages

- Results **put into question** the measurement quality of web tracking measurements
  - Some concepts are measures very accurately: **communication and video streaming**
  - While other **entertainment**

Even if surprising, some of these results make logical sense when we think about the theory of the potential error causes of web tracking data!

Surveys, on the other hand, perform acceptably well. They also struggle more with news, but their quality is never below .50 and generally around .70 (agrees w/ Alwin)

Thanks!

Oriol J. Bosch | Postdoctoral Researcher, University of Oxford

oriol.bosch-jover@demography.ox.ac

orioljbosch

https://orioljbosch.com/

European Research Council
Established by the European Commission

What about the consumption of general news?

Web tracking

- Reliability: 0.38
- Validity: 0.94
- Measurement quality: 0.36

Survey

- Reliability: 0.96
- Validity: 0.77
- Measurement quality: 0.71

What about undercoverage?

Average measurement quality for fully covered sample

- News: 0.30
- Communication: 0.81
- Entertainment: 0.53

Survey

- News: 0.22
- Communication: 0.80
- Entertainment: 0.49