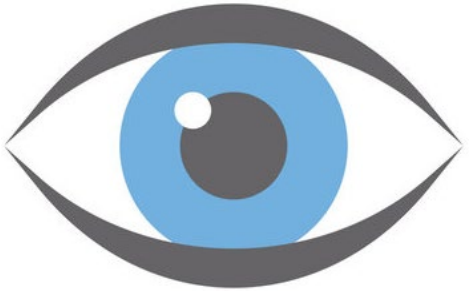
Beyond Clicks and Text: Integrating new types of data into web surveys

CS3 Meeting for Computational Survey and Social Science

28 January 2026

Melanie Revilla | RECSM-UPF

Acknowledgments:
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 849165).

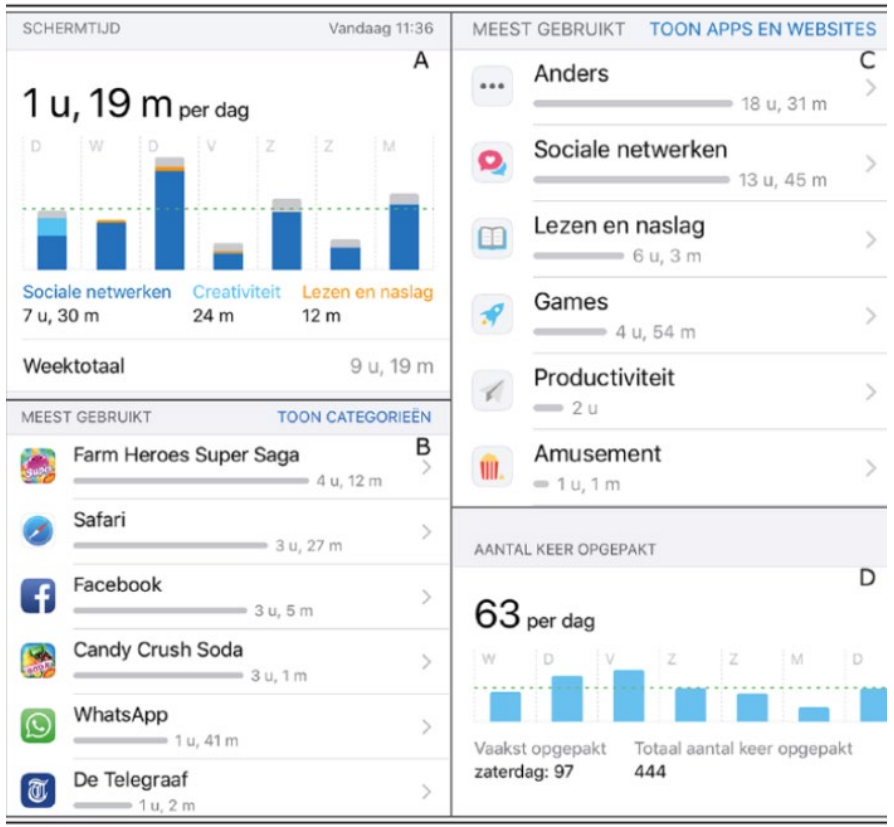


Focus on visual data

USE OF VISUAL DATA

Visual data have been used to study different topics

Screen-time

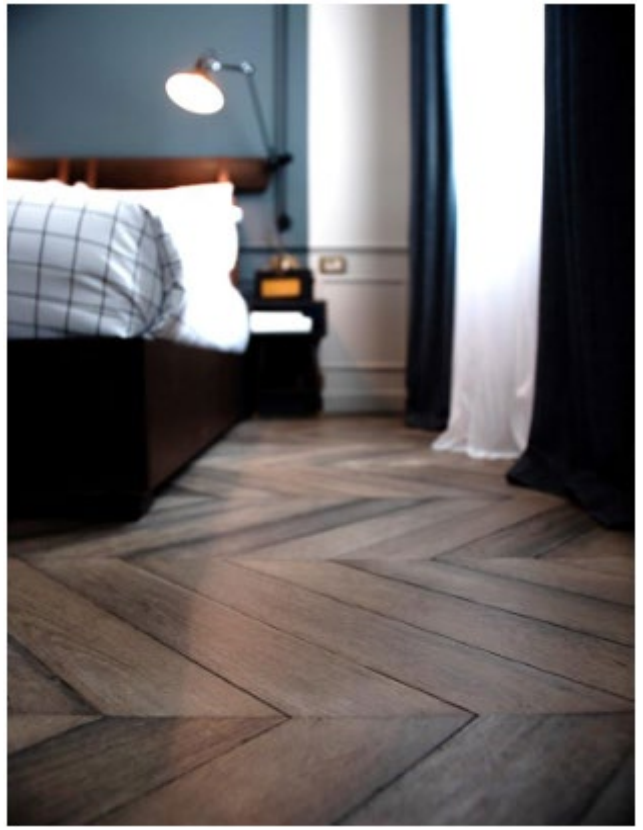


Ohme et al. (2021)

Heating systems

Ilic et al. (2022)

Bedroom flooring



Slavec (2024)

WHY?

Main expected benefits & challenges (Revilla, 2022)

web data opp

Researchers

Reduce some of the issues related to measurement errors

Social desirability bias

	EUR
Banana	1,44 A
1,148 kg x 1,25 EUR/kg	
Freshona/Espinacas	1,15 A
Vemondo/Bebida soja 0%	1,60 B
2 x 0,80	
Vemondo/Tofu ecológico	0,95 B
Mandarina Ebre	2,79 A
Dentalux/Crema dental	0,95 C
Chef Select/Trio de humm	2,19 B
Edulis/Ensalada dúo	1,15 A
Alesto/Mezcla frutos sec	1,89 B
Floralys/Servill 2capas	0,95 C
Favorina/Huevos chocolat	1,49 B
Champiñón	0,65 A
Huevos L suelo	1,79 A
Floralys/Papel higiénico	2,55 C
Total	21,54

Benefits

Participants

WHY?

Main expected benefits & challenges (Revilla, 2022)

web data opp

Researchers

Reduce some of the issues related to measurement errors

Obtain new insights

Information respondents do not know

Benefits

Participants

WHY?

Main expected benefits & challenges (Revilla, 2022)

web data opp

Benefits

Researchers

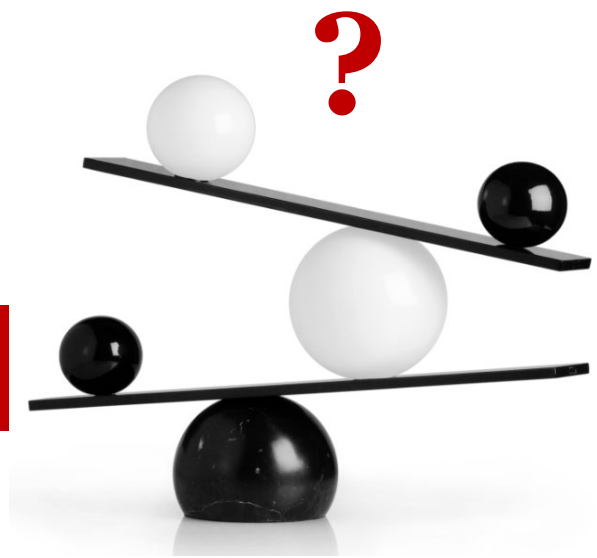
Reduce some of the issues related to measurement errors

Obtain new insights

Participants

Reduced time/effort dedicated to provide information

More enjoyable



Research needed!

Challenges

Researchers

Selection bias

Data protection & ethical issues

Different types of errors

Participants

Privacy issues

Loss of control

New skills needed

COLLECTING AND PROCESSING IMAGES (Iglesias et al., 2024)

An 8-step procedure to guide researchers

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

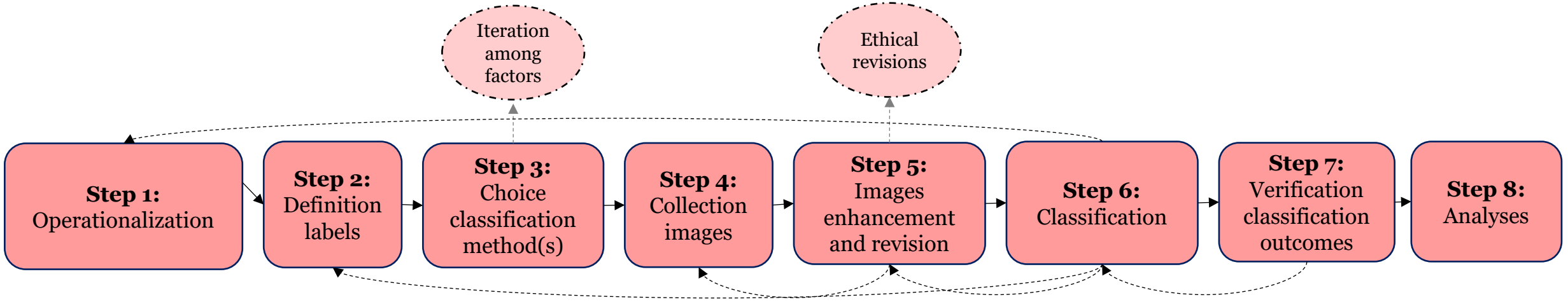
A practical guide to (successfully) collect and process images through online surveys

Patricia A. Iglesias [a,*], Carlos Ochoa [a], Melanie Revilla [b]

[a] Research and Expertise Centre for Survey Methodology (RECSM), Department of Political and Social Sciences, Universitat Pompeu Fabra, Barcelona, Ramon Trias Fargas, 25-27, 08005, Barcelona, Spain
[b] Institut Barcelona d'Estudis Internacionals (IBEI), C/ de Ramon Trias Fargas, 25-27, 08005, Barcelona, Spain

A B S T R A C T

Asking online survey respondents to share images is a practice that has gained notoriety recently. Although this collecting strategy may offer many advantages, it requires researchers to know how to operationalize, collect, process, and analyze this type of data, which is not yet an extended expertise among survey practitioners. This paper aims to guide researchers inexperienced in image analysis by presenting the main steps involved in the process of using images as a new data source: 1) operationalization, 2) definition of the labels, 3) choice of the most suitable classification method(s), 4) collection, 5) enhancement, and 6) classification of the images, 7) verification of the classification outcomes, and 8) data analysis. Following this eight-step process can help practitioners assess whether image collection is appropriate for their research problem and, if so, plan their image-based research, by providing them with the key considerations and decisions to address throughout their implementation.



Case study 1: books-at-home

CASE STUDY 1: BOOKS-AT-HOME

Relevance

- Number of books often used as indicator of cultural or economic capital
  - But people do not know how many books they have
  - Social desirability bias expected → over-reporting
  - Kind of books also matter (cooking vs history books)

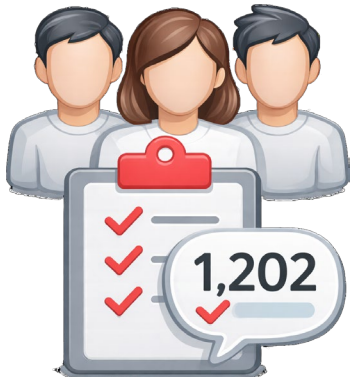
- Asking for photos of the books has the potential to provide:
  - More accurate information about the number of books
  - Extra information (kind of books, language, storage, etc.)

A *picture* is **worth** a **thousand** words

CASE STUDY 1: BOOKS-AT-HOME

Data collection

- Online survey
- Smartphone or tablet only
- June 2023



1,202 respondents

- Target population: parents of children in 1st, 3rd, or 5th year of primary school
- Quotas for gender, age, and level of education

CASE STUDY 1: BOOKS-AT-HOME

Variables of interest

Number of books	Language of books	Storage
Illiterate children Literate children General audience	Spanish Other official language Other language	Shelves Tables Closets Other?

CASE STUDY 1: BOOKS-AT-HOME

Experimental design

4 groups

Group: ***Choice*** *(n=305)*	*Text* or *Images*
Group: ***Text-TextPlus*** *(n=304)*	*Text* → *TextPlus*
Group: ***TextPlus-Images*** *(n=332)*	*TextPlus* → *Images*
Group: ***Images-Text*** *(n=329)*	*Images* → *Text*

CASE STUDY 1: BOOKS-AT-HOME

2 papers using these data

methods, data, analyses | Vol. 19(2), 2025, pp. 46-86 DOI: 10.12758/mda.2024.07

Preferences, Participation, and Evaluation of Answering Questions About the Books Participants Have at Home Through Conventional and Image-Based Formats

Patricia A. Iglesias

Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra

Abstract

The collection of photos through online surveys has emerged as a valuable research tool given the growing use of smartphones, which have facilitated the capture and share of photos. However, gaps persist in understanding respondents' involvement in these tasks when asked to perform them in an online survey. Existing literature lacks insights into participants' preferences, their assessment of questions asking for photos, and how their characteristics might impact their participation in such queries. This paper addresses these gaps, while also comparing how image-based formats compare to conventional ones. Conducted among 1,270 parents living with children in primary school of an opt-in panel in Spain, the mobile online survey implemented in this study revealed a preference for conventional questions, and higher participation in that format than in the image-based one. Respondents able to choose their response format and preferring images presented higher participation rates than those without a choice. While both formats were perceived as equally easy, participants using conventional formats liked the questions better than those answering through photos. Finally, age, being female, having a tertiary education degree, and using the camera at least once a week positively impacted the participation in image-based questions, whereas comfort with new technologies increased the likelihood of liking this format. This study not only fills critical gaps in the literature but also sheds light on the complexities of asking for photos in online surveys.

Quality & Quantity
https://doi.org/10.1007/s11135-025-02519-7

Unlocking insights: assessing the quality of conventional and image-based responses on books at home in an online mobile survey

Patricia A. Iglesias[1,2]

Received: 29 November 2024 / Accepted: 25 November 2025

Abstract

Despite growing interest in collecting photos within online surveys, little is known about the quality of visual data and its comparison with data obtained through conventional requests. To address this gap, a self-administered online mobile survey targeting parents of children attending primary school in Spain was conducted through the Netquest opt-in panel in 2023. The survey gathered information about books in respondents' homes through photos and conventional questions. First, a review of previous research using conventional questions, photos, and other emerging data types was conducted to identify indicators suitable to evaluate the quality of the information about books at home collected through conventional and image-based formats. Second, most of these indicators to measure quality were estimated. Results reveal important measurement errors in conventional questions, while photos submitted by respondents are generally in line and can be classified. However, concrete information of interest about the books, such as the intended audience or languages, is often difficult to extract from photos. When comparing quality, conventional answers provide more information about the items asked than photos, but photos have the potential to provide additional insights, such as book titles. Overall, while collecting and analyzing photos sent through surveys presents challenges, their integration into surveys offers unique opportunities to enrich data collection methods.

CASE STUDY 1: BOOKS-AT-HOME

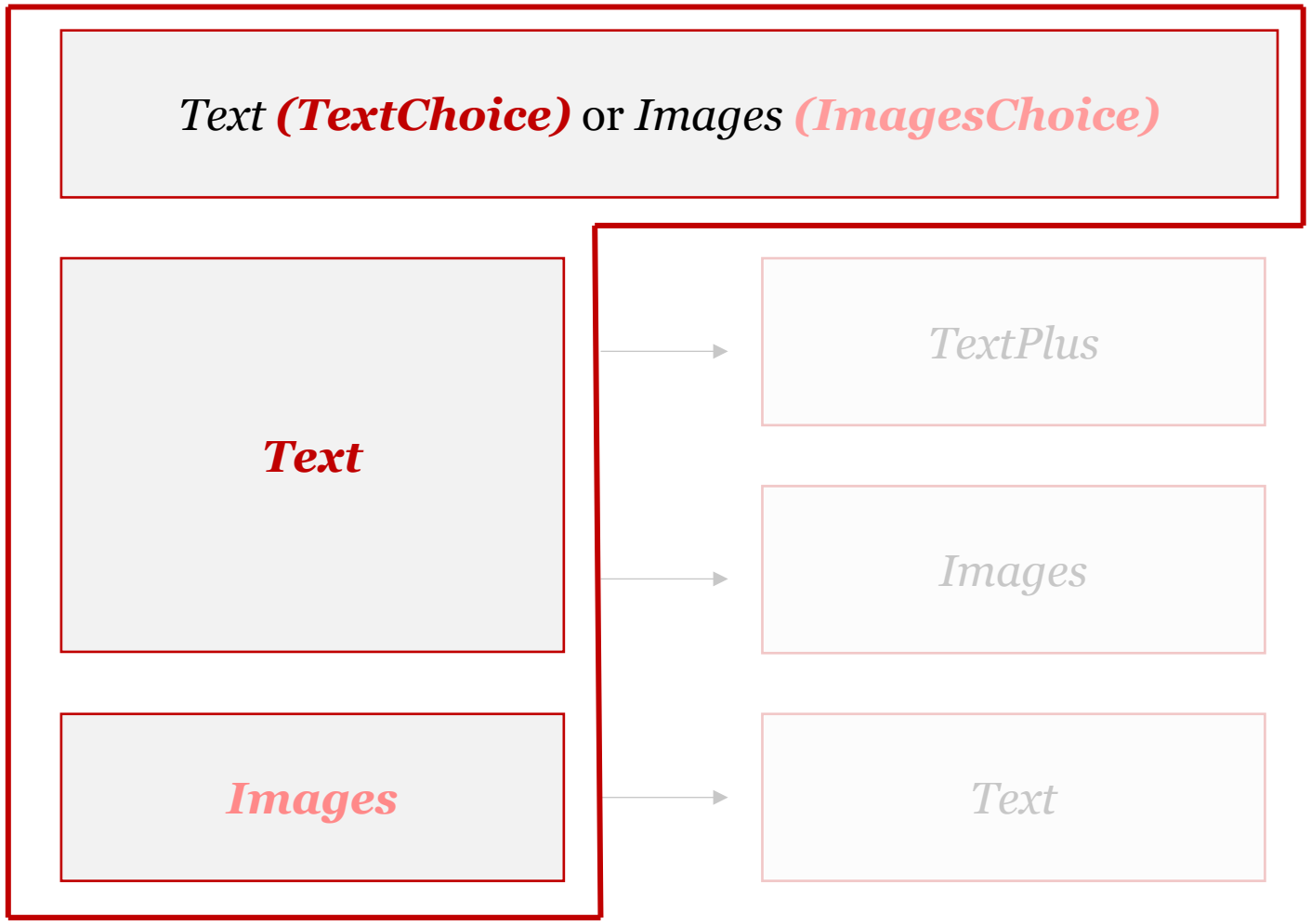
Preferences, participation, evaluation (Iglesias, 2024)

In this paper, Text and TextPlus combined + focus on time 1

Group: ***Choice*** *(n=305)*

Group: ***Text*** *(n=636)*

Group: ***Images-Text*** *(n=329)*



CASE STUDY 1: BOOKS-AT-HOME

Preferences, participation, evaluation (Iglesias, 2024)

Main results

Preferences	Participation	Evaluation
Notable preference for conventional question formats Only 4% choosing the photos when given a choice	Higher for conventional questions (80%) compared to requests for photos (40%)	Participants perceived both formats as equally easy But those answering through conventional questions liked it better

Photos might still not be ready to replace conventional questions (at least for this topic), but they could serve as valuable complements

CASE STUDY 1: BOOKS-AT-HOME

2 papers using these data

methods, data, analyses | Vol. 19(2), 2025, pp. 46-86 DOI: 10.12758/mda.2024.07

Preferences, Participation, and Evaluation of Answering Questions About the Books Participants Have at Home Through Conventional and Image-Based Formats

Patricia A. Iglesias

Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra

Abstract

The collection of photos through online surveys has emerged as a valuable research tool given the growing use of smartphones, which have facilitated the capture and share of photos. However, gaps persist in understanding respondents' involvement in these tasks when asked to perform them in an online survey. Existing literature lacks insights into participants' preferences, their assessment of questions asking for photos, and how their characteristics might impact their participation in such queries. This paper addresses these gaps, while also comparing how image-based formats compare to conventional ones. Conducted among 1,270 parents living with children in primary school of an opt-in panel in Spain, the mobile online survey implemented in this study revealed a preference for conventional questions, and higher participation in that format than in the image-based one. Respondents able to choose their response format and preferring images presented higher participation rates than those without a choice. While both formats were perceived as equally easy, participants using conventional formats liked the questions better than those answering through photos. Finally, age, being female, having a tertiary education degree, and using the camera at least once a week positively impacted the participation in image-based questions, whereas comfort with new technologies increased the likelihood of liking this format. This study not only fills critical gaps in the literature but also sheds light on the complexities of asking for photos in online surveys.

Quality & Quantity
https://doi.org/10.1007/s11135-025-02519-7

Unlocking insights: assessing the quality of conventional and image-based responses on books at home in an online mobile survey

Patricia A. Iglesias[1,2]

Received: 29 November 2024 / Accepted: 25 November 2025

Abstract

Despite growing interest in collecting photos within online surveys, little is known about the quality of visual data and its comparison with data obtained through conventional requests. To address this gap, a self-administered online mobile survey targeting parents of children attending primary school in Spain was conducted through the Netquest opt-in panel in 2023. The survey gathered information about books in respondents' homes through photos and conventional questions. First, a review of previous research using conventional questions, photos, and other emerging data types was conducted to identify indicators suitable to evaluate the quality of the information about books at home collected through conventional and image-based formats. Second, most of these indicators to measure quality were estimated. Results reveal important measurement errors in conventional questions, while photos submitted by respondents are generally in line and can be classified. However, concrete information of interest about the books, such as the intended audience or languages, is often difficult to extract from photos. When comparing quality, conventional answers provide more information about the items asked than photos, but photos have the potential to provide additional insights, such as book titles. Overall, while collecting and analyzing photos sent through surveys presents challenges, their integration into surveys offers unique opportunities to enrich data collection methods.

CASE STUDY 1: BOOKS-AT-HOME

Data quality (Iglesias, 2026)

Different indicators of data quality used in the literature, but most of them cannot be applied to visual data (e.g., rounding, straightlining…)

Need to think further about how to evaluate quality in the case of visual data, and how to compare data quality across conventional and photo requests

Iglesias (2026) elaborated a list of 17 indicators to assess data quality of conventional and photo requests

- 4 for conventional questions; 6 for photos and 7 for both formats

Then, she estimated these indicators using the data on books at home

CASE STUDY 1: BOOKS-AT-HOME

Data quality (Iglesias, 2026)

- However, before estimating data quality, necessary to classify the photos
  - Human classification with 2 trained classifiers sharing the work
    - Overlap of 100 images (out of 723) to help improving classification
  - Detailed guidelines developed

Important challenges for classifying

- Overlap

CASE STUDY 1: BOOKS-AT-HOME

Data quality (Iglesias, 2026)

- However, before estimating data quality, necessary to classify the photos
  - Human classification with 2 trained classifiers sharing the work
    - Overlap of 100 images (out of 723) to help improving classification
  - Detailed guidelines developed

Important challenges for classifying

- Overlap
- Hidden books
- Books or not?
- Etc.

CASE STUDY 1: BOOKS-AT-HOME

Data quality (Iglesias, 2026)

To study data quality, Choice group excluded

Group: ***Choice*** *(n=305)*	*Text* or *Images*		
Group: ***Text-TextPlus*** *(n=304)*	*Text*	→	*TextPlus*
Group: ***TextPlus-Images*** *(n=332)*	*TextPlus*	→	*Images*
Group: ***Images-Text*** *(n=329)*	*Images*	→	*Text*

Data quality (Iglesias, 2026)

Main results

- Conventional formats exhibited notable shortcomings in terms of data quality
- Photos generally met the criteria for visual quality and were in line with what was asked (99%)
- However, only in 4% of cases, all the information of interest could be obtained
  - Problem of classification, especially of the type of books (for illiterate vs literate children) and languages
- Important discrepancies between formats, especially for the number and categorization of books

Case study 2:

remote work

CASE STUDY 2: REMOTE WORK

Relevance (of topic)

Increase in remote work + Working conditions affect productivity and physical & mental well-being = Crucial to study whether people have **acceptable remote working conditions**

Problems: social desirability + high burden

↓

Solution?
Photos can provide detailed and objective information about the remote workstations with limited effort

CASE STUDY 2: REMOTE WORK

Relevance (for methodology field)

Participation still a key challenge

Need to test different strategies to improve participation when requesting visual data

This study tests the effect on participation + data quality of:

Follow-up

Extra incentive

Reminder email

Main survey

Reminder

CASE STUDY 2: REMOTE WORK

Data collection

- Online survey
- Smartphone or tablet only
- December 2025

2,327 eligible respondents started

- Target population: adults aged 18–65 who had worked from home for 7+ hours per week on average during the previous 2 months
- Soft quotas for gender & age

CASE STUDY 2: REMOTE WORK

Photos requested

Photo 1

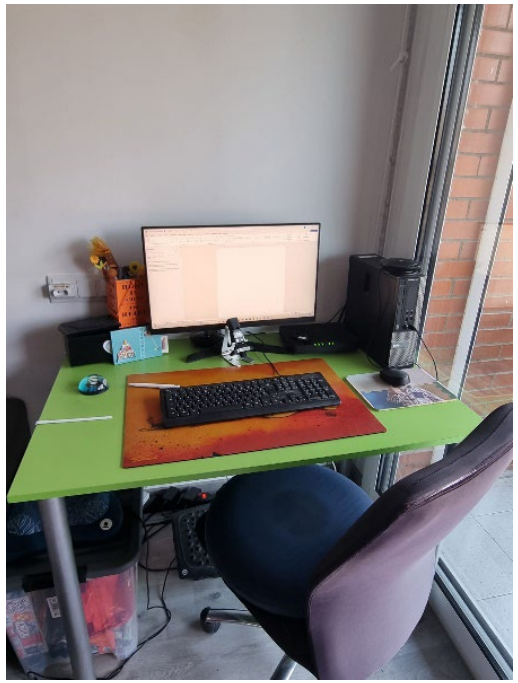
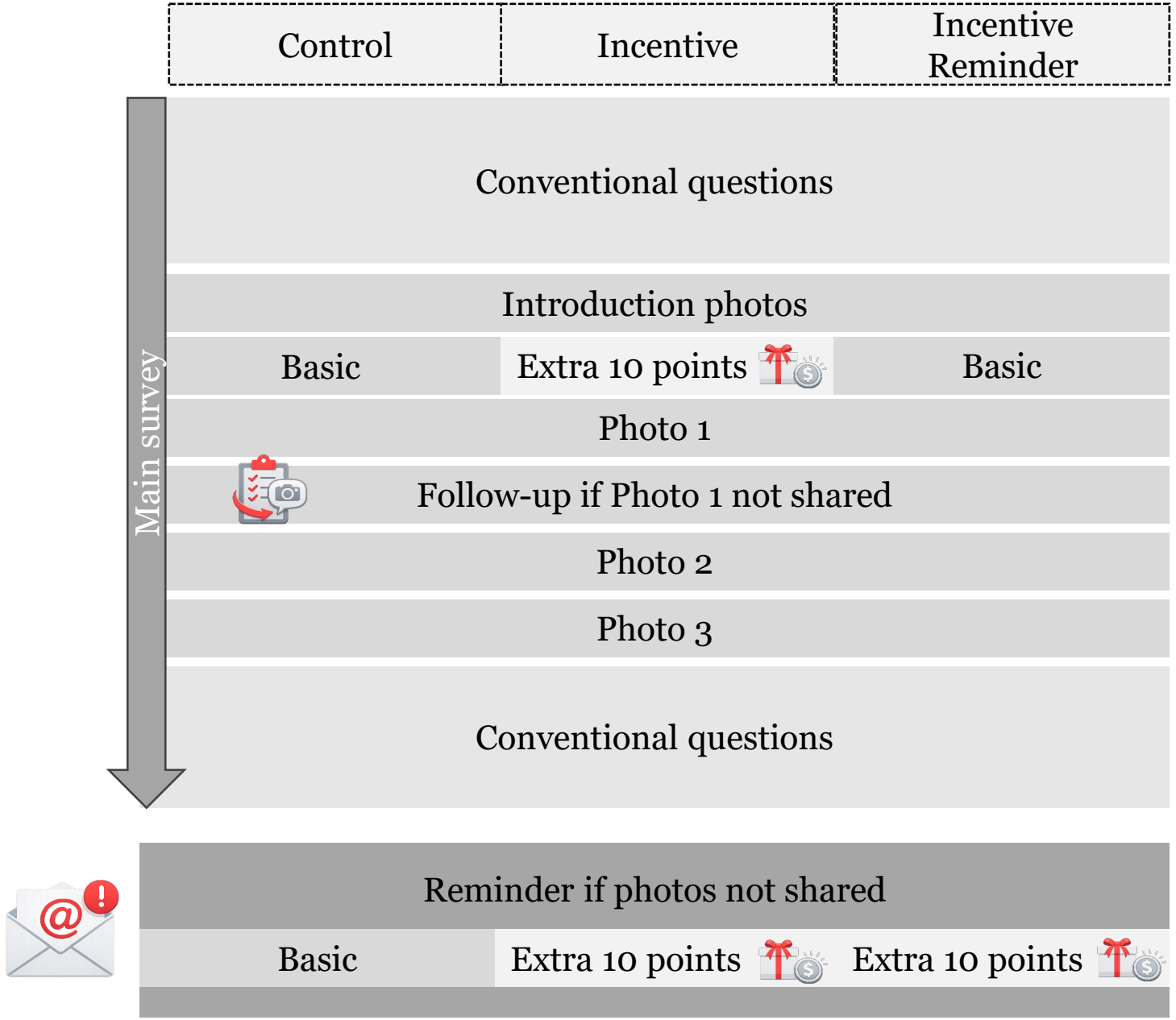


Photo 2

Photo 3

CASE STUDY 2: REMOTE WORK

Study design



CASE STUDY 2: REMOTE WORK

First very initial results (Padilla & Revilla, in progress)

Focusing on Photo 1

2,327 Started survey + eligible

1,354 Assigned to an experimental group*

742 Shared Photo 1

124

58

54.8%* **63.9%*** **68.2%***

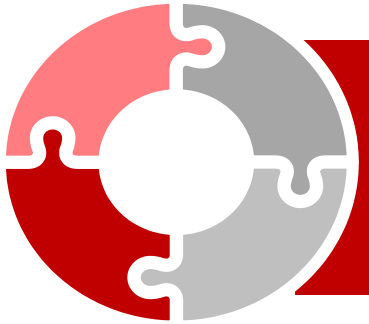
+9.1

+4.3

- Follow-up really useful
- Reminder also improves
- Both strategies should be considered (low cost, potential gain)
- Further analyses needed: effect incentives, validity photos, etc.

web data opp

Conclusions

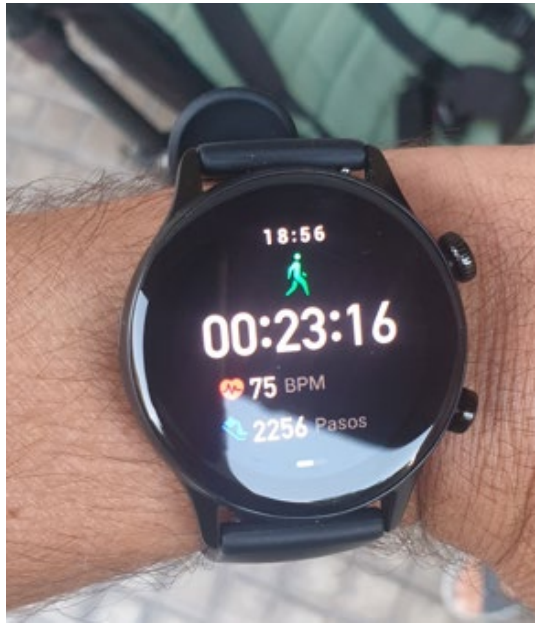


Starting is difficult,
finishing is way harder

Clear potential

- Visual data have the potential to provide very rich information
  - Including about aspects respondents do not know
  - And things that are very hard to express in words

- Many possible applications



But challenges

- However, collecting visual data through web surveys is **not easy**
  - Researchers should carefully weigh the pros and cons before deciding to collect images
  - Likely to become easier in the future
  - But to this day, still a lot of challenges

- Some research nevertheless started to shed some light and can help researchers who want to collect visual data for the first time
  - 8-step process of Iglesias et al. (2024) is a good starting point
  - But each data collection has its **specificities** so important to carefully consider each new application

- Still a lot to be done: methodological + substantive research

Thanks!

Questions?

Melanie Revilla | RECSM-UPF

melanie.revilla@upf.edu

https://www.upf.edu/web/webdataopp

Universitat Pompeu Fabra Barcelona

RECSM Research and Expertise Centre for Survey Methodology

webdataopp

REFERENCES

- Iglesias, P.A. (2026). Unlocking insights: assessing the quality of conventional and image-based responses on books at home in an online mobile survey. Quality & Quantity. https://doi.org/10.1007/s11135-025-02519-7
- Iglesias, P. A. (2024). Preferences, Participation, and Evaluation of Answering Questions About the Books Participants Have at Home Through Conventional and Image-Based Formats. methods, data, analyses, 1-36. https://doi.org/10.12758/mda.2024.07
- Iglesias, P.A., Ochoa, C., & Revilla, M. (2024). A practical guide to (successfully) collect and process images through online surveys. Social Sciences & Humanities Open, 9, 100792. https://doi.org/10.1016/j.ssaho.2023.100792.
- Ilic, G., Lugtig, P., Schouten, B., Streefkerk, M., Mulder, J., Kumar, P., & Höcük, S. (2022). Pictures Instead of Survey Questions: An Experimental Investigation of the Feasibility of Using Pictures in a Housing Survey, Journal of the Royal Statistical Society Series A: Statistics in Society, 185 (Supplement_2): S437–S460, https://doi.org/10.1111/rssa.12960
- Ohme, J., Araujo, T., de Vreese, C.H., & Piotrowski, J.T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. Mobile Media & Communication, 9(2), 293-313. https://doi.org/10.1177/2050157920959106
- Revilla, M. (2022). How to enhance web survey data using metered, geolocation, visual and voice data?. *Survey Research Methods*, 16(1): 1-12. https://doi.org/10.18148/srm/2022.v16i1.8013
- Slavec, A.(2024). Visual data in web surveys: the case of bedroom flooring. WEB DATA OPP workshop (March 18-19). Presentation available at: https://www.upf.edu/web/webdataopp/workshop-2024