

# Measuring and comparing the quality of information collected through photos and conventional questions: the case of the books at home.

*Doctoral workshop, 11-12 March*

Patricia A. Iglesias

# Introduction

- Photos, either captured with a mobile device camera or already stored in such devices, could be sent in online surveys as a way to **replace/complement conventional survey questions**.
- Potential benefits of photos: reduce respondents' burden and mistakes, collect information respondents do not know about, decrease social desirability bias, and provide more/new insights (Revilla, 2022).
- Most research on photo collection through online surveys has focused on willingness to participate and participation. However, **little has been said about the quality of that information and how it compares with data obtained through conventional requests**.
- In this paper, I will focus on the **quality** of the information that respondents of an online survey provided through images and/or conventional questions about **the books at home**.

## Why books at home?

- The **number of books at home** is a variable broadly present in social science surveys. However, it presents some limitations:
  - It is very likely that respondents **do not know the exact number of books** at their household → at best, such number can be estimated.

## Why books at home?

- The **number of books at home** is a variable broadly present in social science surveys. However, it presents some limitations:
  - It is very likely that respondents **do not know the exact number of books** at their household → at best, such number can be estimated.
  - Social desirability bias might play a role in **over reporting the estimated number of books**.

# Why books at home?

- The **number of books at home** is a variable broadly present in social science surveys. However, it presents some limitations:
  - It is very likely that respondents **do not know the exact number of books** at their household → at best, such number can be estimated.
  - Social desirability bias might play a role in **over reporting the estimated number of books**.
  - The response categories for this question are usually composed of broad/different intervals.

Response category
"10 books or less"
"11 to 25 books"
"26 to 100 books"
"101 to 200 books"
"201 to 500 books"
"More than 500 books"

(Sieben & Lechner,  
2019)

Número de libros en el hogar
0-10 libros
11-25 libros
26-100 libros
101-200 libros
Más de 200 libros

(Gil Flores, 2011)

0-10 book(s)
11-25 books
26-100 books
101-200 books
201-500 books
More than 500 books

(Güre et al., 2023)

None
Less than 10
10 to 20
21 to 30
More than 30 books

(McNally et al.,  
2023)

## Why books at home?

- The **number of books at home** is a variable broadly present in social science surveys. However, it presents some limitations:
  - It is very likely that respondents **do not know the exact number of books** at their household → at best, such number can be estimated.
  - Social desirability bias might play a role in **over reporting the estimated number of books**.
  - The response categories for this question are usually composed of broad/different intervals.
  - As it is a variable used as proxy of cultural and socioeconomic capital, **the number of books alone might not be sufficient** → type of books (from the title), which would be too burdensome for respondents.
- These limitations could be overcome by photos of the books at home, **potentially increasing the quality of the data and providing new insights**.

# Background: measurement of quality in the literature

## Assessment of the quality of conventional questions

- There are many indicators used for measuring the quality of conventional questions:
  - rounding,
  - item nonresponse,
  - completion time,
  - incoherence of responses,
  - length of answers to open-narrative questions,
  - response distribution,
  - and so on....

# Background: measurement of quality in the literature

## **Assessment of the quality of photos collected through online surveys**

- Photos in line with what is being asked (Bosch et al., 2019, Ilic et al. 2019) → that the content of the photos relates with what is being asked.
- Number of photos (Ilic et al., 2021).
- Photos allowing the classification of the information (Ilic et al., 2021).

**Not much has been elaborated to assess the quality of photos or to compare that quality with the one of conventional questions → research gap.**



# Objective and research questions

**Objective:** To assess the quality of the information about the books at home when provided through conventional questions or through photos.

Number of  
books

Languages of  
the books

Storage of the  
books

# Objective and research questions

**Objective:** To assess the quality of the information about the books at home when provided through conventional questions or through photos.

Number of  
books

Languages of  
the books

Storage of the  
books

**RQ1:** Which indicators can be used to assess the quality of the information about the books at home collected through conventional and images-based formats?

# Objective and research questions

**Objective:** To assess the quality of the information about the books at home when provided through conventional questions or through photos.

Number of  
books

Languages of  
the books

Storage of the  
books

**RQ1:** Which indicators can be used to assess the quality of the information about the books at home collected through conventional and images-based formats?

**RQ2:** What is the quality of the information about the books at home when collected through conventional and images-based formats?

# Methods and data

## **To collect the information about books:**

- Online survey that could only be answered from a smartphone or tablet.
- Collected through the Netquest online panel in Spain, in June 2023.
- Target population: parents of children in first, third, or fifth year of primary school.
- Quotas for gender, age, and level of education.
- 1,202 respondents completed the survey.

## Methods and data

Participants were asked a) to answer questions about the books they have at home, and/or b) to send photos of such books.

# Methods and data

Participants were asked a) to answer questions about the books they have at home, and/or b) to send photos of such books.

## Conventional questions:

- **the number of books:** four open-ended questions about 1) the total number of books at home, and the number of books 2) for toddlers and children who do not know how to read, 3) for literate children and teenagers, and 4) aimed at a general audience.
- **language:** three open-ended questions asking for the percentage of books 1) in Spanish, 2) in one of the three co-official languages in Spain (Catalan, Galician, and Euskera), and 3) in other languages.
- **storage:** four radio-button questions asking whether books are stored 1) in shelves, 2) inside closets or drawers, 3) in center, coffee, or night tables or over a desk, and 4) in other places.

# Methods and data

Participants were asked a) to answer questions about the books they have at home, and/or b) to send photos of such books.

## Conventional questions:

- **the number of books:** four open-ended questions about 1) the total number of books at home, and the number of books 2) for toddlers and children who do not know how to read, 3) for literate children and teenagers, and 4) aimed at a general audience.
- **language:** three open-ended questions asking for the percentage of books 1) in Spanish, 2) in one of the three co-official languages in Spain (Catalan, Galician, and Euskera), and 3) in other languages.
- **storage:** four radio-button questions asking whether books are stored 1) in shelves, 2) inside closets or drawers, 3) in center, coffee, or night tables or over a desk, and 4) in other places.

# Methods and data

Participants were asked a) to answer questions about the books they have at home, and/or b) to send photos of such books.

## Conventional questions:

- **the number of books:** four open-ended questions about 1) the total number of books at home, and the number of books 2) for toddlers and children who do not know how to read, 3) for literate children and teenagers, and 4) aimed at a general audience.
- **language:** three open-ended questions asking for the percentage of books 1) in Spanish, 2) in one of the three co-official languages in Spain (Catalan, Galician, and Euskera), and 3) in other languages.
- **storage:** four radio-button questions asking whether books are stored 1) in shelves, 2) inside closets or drawers, 3) in center, coffee, or night tables or over a desk, and 4) in other places.



# Methods and data

Participants were asked a) to answer questions about the books they have at home, and/or b) to send photos of such books.

## Conventional questions:

- **the number of books:** four open-ended questions about 1) the total number of books at home, and the number of books 2) for toddlers and children who do not know how to read, 3) for literate children and teenagers, and 4) aimed at a general audience.
- **language:** three open-ended questions asking for the percentage of books 1) in Spanish, 2) in one of the three co-official languages in Spain (Catalan, Galician, and Euskera), and 3) in other languages.
- **storage:** four radio-button questions asking whether books are stored 1) in shelves, 2) inside closets or drawers, 3) in center, coffee, or night tables or over a desk, and 4) in other places.

## Images-based question:

Sending photos of all the books in the household.

The information of interest could be classified from the photos.

# Methods and data

The conventional questions were collected through two methods:

## **Text**

Asked for the number, language, and storage of books by using conventional questions.

## **TextPlus**

Similar to **Text**, but with a visual example for the number-of-books question.

## Methods and data

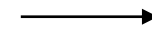
**Message for group TextPlus:** *To help you estimate the total number of books that you have in your main residence, please, look at the examples below: you can see that a 74 centimeters long shelf can contain from around 30 to almost 80 books, depending on the thickness of the books.*



# Methods and data

Group **Text-TextPlus**  
(*n=301*)

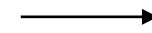
Text



TextPlus

Group **TextPlus-Images**  
(*n=300*)

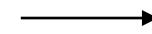
TextPlus



Images

Group **Images-Text**  
(*n=300*)

Images



Text

# Methods and data

To answer **RQ1** (indicators that can be used to assess the quality of the information):

## Conventional questions → applying what already exists in the literature

- Rounding (# of books and % of languages)
- Inconsistencies between the stated total # of books and the sum of the three categories.
- Inconsistencies in the sum of languages not being 100%.
- Inconsistencies between Text and TextPlus.
- Proportion of DK in # of books.

# Methods and data

To answer **RQ1** (indicators that can be used to assess the quality of the information):

## Conventional questions → applying what already exists in the literature

- Rounding (# of books and % of languages)
- Inconsistencies between the stated total # of books and the sum of the three categories.
- Inconsistencies in the sum of languages not being 100%.
- Inconsistencies between Text and TextPlus.
- Proportion of DK in # of books.

## Photos → literature + new indicators

- Photos in line (contain books or not)
- Classification potential:
  - Books can be counted
  - Languages can be identified
  - Storage can be identified
- Number of photos
- Prevalence of problems (technical issues, not being able to capture all the photos, among others)

# Methods and data

To answer **RQ1** (indicators that can be used to assess the quality of the information):

## Indicators for both types of data:

- Full response: percentage of substantive answers providing information for all the dimensions.
- Comparison of the reported number of books between formats based on the order of the questions → no true value, but one could expect:
  - Over report with the conventional format → social desirability bias
  - Under report with the images → inconvenience of making all the photos

## Methods and data

To answer **RQ2** (what is the quality of the information):

Applying the previous indicators to the information collected.



# What is pending

- Define the exact calculations for each indicator.
- Explore further indicators for measuring the quality of photos → new indicators should be created?
- Define how to differentiate (if needed) the quality of photos at the photo and at the respondent level.
- See quality measures specific for the number of books at home (for conventional questions).
- Assess and compare the quality.

# Thanks!

## *Questions?*

Patricia Iglesias



[patricia.iglesias@upf.edu](mailto:patricia.iglesias@upf.edu)



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