SMART SURVEYS AND THE DIGITAL DIVIDE IN ITALY: THE INFLUENCE OF SOCIO-DEMOGRAPHIC CHARACTERISTICS ON THE SMART TASKS' PERFORMANCE[1]

Monica Perez, Barbara D'Amen,
Barbara Maria Rosa Lorè, Alessandra Nuccitelli

Abstract. Smart surveys combine data provided by the respondent, either in a traditional way via a web questionnaire or through smart instruments (active mode), with data collected by sensors embedded in smart devices (passive mode). This approach reduces the burden on respondents and improves the quality and timeliness of the results, especially in complex surveys. Within the framework of the European Statistical System, Istat participated in the Smart Surveys Implementation project (2023-2025), in partnership with other national statistical institutes and some European universities. As part of the project, an experimental survey named "New Ways of Measuring" was conducted in Italy, the Netherlands, and Slovenia in a harmonised manner with the aim of assessing the respondents' attitudes towards the new ways of collecting data in smart surveys.
The purpose of this study is to investigate how socio-demographic characteristics influence the performance of smart tasks in the Italian context. The findings provide valuable insights into the relationship between innovations in survey techniques and digital inequality.

1. Introduction

Over the past decade, official statistics have undergone a profound transformation driven by digitalisation and the increasing use of smart devices and mobile technologies. These developments have paved the way for innovative data collection methods that go beyond traditional questionnaire-based surveys. Among these, smart surveys, which integrate self-reported data with sensor-generated information, have emerged as a promising tool. In this approach, respondents may be asked not only to answer questions via web questionnaires (active data), but also to consent to the passive collection of behavioural or contextual information, such as geolocation, physical activity, or purchase transactions recorded through their smartphones or wearable devices (Struminskaya *et al.*, 2020). By leveraging both active and passive data sources, smart surveys aim to reduce the response burden, improve the timeliness of data collection, and enhance its accuracy, particularly in domains where

[1] This work is the result of the joint effort by the authors. However the Sections can be attributed as follows: Sections 1 and 4 to Barbara M.R. Lorè, Section 2 to Monica Perez, Sections 3 and 6 to Alessandra Nuccitelli, Section 5 to Barbara D'Amen.

continuous or real-time measurement is valuable. However, the implementation of smart surveys introduces new methodological and ethical challenges. One of the most pressing issues is the digital divide, a multidimensional concept encompassing disparities not only in the access to technology, but also in skills, usage patterns, and trust in digital environments (Mossberger *et al.*, 2003). In the context of smart surveys, this divide can have a significant impact on the feasibility of data collection and the representativeness of final data. Individuals with lower digital literacy or fewer technological resources may be less likely to engage with smart tasks or grant consent to data sharing.

Research has shown that socio-demographic characteristics – especially age, education, income, and geographical area – are significant predictors of digital competence and device usage (Van Deursen and Van Dijk, 2014). These factors are also associated with varying levels of trust in digital systems – including concerns about data privacy and security – that may further hinder participation in smart data collection (Keusch *et al*., 2023). Therefore, it is essential to understand how these characteristics affect smart task performance in order to design equitable and effective data collection strategies and to mitigate the non-response bias introduced by the use of the new survey tools.

The present study investigates how socio-demographic characteristics influence respondents' propensity to perform smart tasks using data from an experimental survey named New Ways of Measuring (NWM). This survey was carried out in Italy by the Italian National Statistical Institute (Istat) as part of the Smart Surveys Implementation (SSI) project (2023-2025), funded by Eurostat[2].

This paper is organised as follows. Section 2 introduces some aspects of the NWM survey, with a focus on the questionnaires and the smart tasks that respondents were asked to complete. Section 3 reports the main findings from a logistic regression model, identifying the factors that mostly affect the propensity to perform these tasks. In Section 4, some descriptive analyses investigate the issues of the digital vulnerability and unequal engagement in smart actions, while in Section 5 a qualitative analysis of the open-ended responses explores the reasons why respondents did not perform the smart tasks. Finally, some conclusions are drawn (Section 6).

[2] Grant Agreement Project n. 101119594 - SSI; Call topic: SMP-ESS-2022-TrustedSmartSurveys-H7-6257.

2. The New Ways of Measuring survey

Under the European Statistical System (ESS), Istat participated in the SSI project (2023-2025) in partnership with other National Statistical Institutes and some European universities. The project aimed to test and implement smart solutions for the collection of data relating to the Time Use Survey and Household Budget Survey in order to relieve the statistical burden deriving from the requirement for respondents to keep detailed diaries. As part of the project, the NWM survey was conducted in Italy, the Netherlands, and Slovenia in a harmonised manner with the aim of assessing the respondents' attitudes towards the new ways of collecting data in smart surveys (Lunardelli *et al.*, 2024). The Italian survey was carried out in early 2024 on a sample of 3,667 individuals aged 18 or older, randomly selected from the population registers of 131 municipalities. Each person was asked to complete two questionnaires sequentially: first a paper NWM-General questionnaire and then an online NWM-Smart questionnaire (henceforth, NWM-General and NWM-Smart will be abbreviated to NWM-G and NWM-S, respectively).

The NWM-G questionnaire was designed to reach also those who do not have smart devices and/or have limited skills in using them. The aim of this questionnaire was to find out how individuals would like to be invited to participate in a statistical survey, the mode and the device through which they would prefer to participate, their habits regarding the use of smart devices (smartphones or tablets), and their opinions about the use of smart devices for data collection. Individuals were also asked about their hypothetical willingness to perform smart tasks, their concerns regarding the theft or misuse of data provided via smart devices, and the measures they consider essential to guarantee the security of their data.

The NWM-S questionnaire, on the other hand, was designed to explore the respondents' actual willingness to share data via smart devices. This questionnaire required the performance of four smart tasks:

- sharing geolocation
- sharing step count
- taking and sharing photos of shopping receipts
- taking and sharing photos of energy meters.

For the NWM-S questionnaire, individuals were advised to use smartphones/tablets (rather than PCs, notebooks, or laptops), as these devices have built-in features (such as webcams) that allow some of the requested actions to be performed more immediately.

The response rate was 68.9% and 25.3% for the NWM-G and NWM-S questionnaires, respectively.

3. The propensity to perform smart tasks

For the respondents to both the NWM-G and NWM-S questionnaires, a logistic regression model is used to analyse the propensity to perform smart tasks. The aim is to investigate the relationship between two exhaustive, non-overlapping categories – "No smart tasks performed" and "At least one smart task performed" – and a set of explanatory variables, either administrative or collected through the NMW-G questionnaire.

A main-effects model is considered, with "No smart tasks performed" taken as the baseline category of the response variable (Hosmer *et al.*, 2012).

The explanatory variables included in the final model are as follows[3]:

- *Number of device types* ("0-1"; "2-3"; "> 3")
- *Household income* per month, in euros ("< 1,900"; "1,900-3,500"; "> 3,500")
- *Concerns about data security*, *i.e.*, concern that the data collected by Istat through smart devices may be stolen or misused by others than Istat ("No", "Yes")
- *Geographical area* ("North"; "Centre, South, and Islands").

Other characteristics are excluded from the final model as they are not statistically significant (p-value > 0.05). In particular, several variables that have been identified as the most influential on the participation in the NMW-S survey – *e.g.*, *Educational level*, *Age*, *Nationality*, *Household size*, *Number of inhabitants in the municipality of residence* – do not seem to be relevant for performing smart tasks.

Table 1 shows the Wald Chi-Square statistic with the associated p-value by explanatory variable (Columns 2 and 3). The ranking relating to the importance of each explanatory variable, as determined by a classification tree[4], is also reported (Column 4). According to this ranking, *Number of device types* and *Household income* are the most influential factors among those considered.

Table 1 – *Wald Chi-Square statistics, p-value, and the ranking determined by the classification tree, by explanatory variable.*

Explanatory variable	Wald Chi-Square	p-value	Ranking
Number of device types	30.4	<0.0001	1
Household income	12.1	0.0024	2
Concerns about data security	8.3	0.0039	4
Geographical area	5.0	0.0258	3

Source: Survey on NWM – year 2024.

[3] The explanatory variables are categorised on the basis of preliminary analyses of their association with the response variable, while also attempting to adopt a parsimonious model.

[4] The classification tree is generated by applying the HPSPLIT procedure of the SAS software, using the Entropy criterion to split the observations and the Cost-complexity pruning method to avoid overfitting (Sas Institute Inc., 2015).

Table 2 shows the maximum likelihood point estimates of the ORs and 95% Wald confidence intervals by regressor[5].

The propensity to perform at least one smart task turns out to increase with both the number of types of devices (OR = 1.68 and 3.19 for "2-3" and "> 3", respectively) and the household income (OR = 1.55 and 1.92 for "1,900-3,500" and "> 3,500", respectively). Compared to the NWM-S respondents who did not engage in any of the smart actions, the smart tasks' performers are much less concerned that the data collected through smart devices might be stolen or misused (OR = 1.83) and are more likely to reside in the North (OR = 1.40).

Table 2 – *Estimated ORs and 95% Wald confidence intervals by regressor.*

Explanatory variable	Contrast	Estimate	CL
Number of device types	2-3 vs *0-1*	1.68	1.17-2.42
	> 3 vs *0-1*	3.19	2.11-4.84
Household income	1,900-3,500 vs < *1,900*	1.55	1.09-2.20
	> 3,500 vs < *1,900*	1.92	1.31-2.80
Concerns about data security	No vs *Yes*	1.83	1.22-2.77
Geographical area	North vs *Centre, South, and Islands*	1.40	1.04-1.89

The reference level of each explanatory variable is in italics.
Source: Survey on NWM – year 2024

4. Digital vulnerability and unequal engagement in smart tasks

A digitally vulnerable subgroup of respondents to both the NWM-G and NWM-S questionnaires is identified on the basis of the characteristics found to be most associated with a lower probability of performing a smart task in the logistic regression model. This *digitally vulnerable subpopulation* includes individuals who reported at least one of the following: (i) using only one type of digital device or none at all; (ii) having a household income below €1,900 per month.

The descriptive analysis that follows compares this subgroup with the general population in terms of engagement with the four smart tasks asked in the NWM-S questionnaire. The aim is to understand to what extent digital disadvantage translates into reduced participation in these tasks and to explore which tasks are more affected by the disparities.

Approximately one in three of the respondents to both questionnaires meets the vulnerability criteria (31.1%). Digitally vulnerable respondents perform worse on all indicators of the digital divide: a lower percentage of users for each type of device,

[5] The maximum likelihood estimates of the model parameters are obtained using the Fisher scoring algorithm.

a lower percentage of frequent users, and a lower level of perceived digital skills (Table 3). Almost two-thirds (63.8%) of vulnerable individuals did not complete any of the four tasks, compared to 52.1% of the total respondents. Only 13.0% of vulnerable individuals completed at least half of the tasks, while this share reaches 17.0% in the general population. These results suggest not only a lower propensity to participate but also a reduced capacity to sustain engagement across multiple tasks within the vulnerable subgroup.

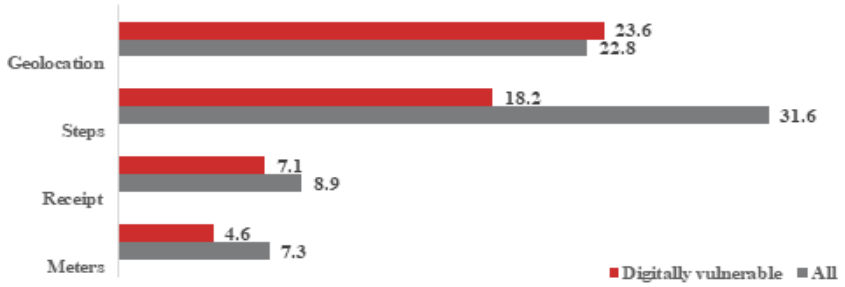
Table 3 – *Digital divide indicators – All and digitally vulnerable respondents.*

Digital divide indicators		All respondents	Vulnerable respondents
Average n. of devices		2.2	0.9
% of users by device:	*Smartphone*	88.3	74.8
	Tablet	41.1	4.0
	Tracker	13.0	1.1
	Watch	23.2	1.9
	Speaker	23.9	2.2
	Electricity meter	13.0	2.0
	Gas meter	11.2	1.1
	Water meter	6.0	0.6
	Air quality monitor	3.9	0.4
% of smartphone daily users		86.6	76.8
Average skill level (1 to 5)		2.9	2.3
Average n. of tasks performed		0.7	0.5

Source: Survey on NWM – year 2024

The four tasks under consideration differ in terms of the tools and skills required for their completion, as well as the burden they impose on respondents. Consequently, the performance on the four tasks varies in relation to both the specifics of each task and the respondents' characteristics (Figure 1).

Figure 1 – *Smart tasks' performance – All and digitally vulnerable respondents (% values).*



Source: Survey on NWM – year 2024

Smart tasks also differ from each other depending on the effort required and the mode of interaction. While some rely on passive or background data sharing (*e.g.*, geolocation, step count), others require an active gesture from the respondent (*e.g.*, taking a picture). In the former case, the information is immediately available to the respondent, whereas in the latter an explicit action is required to retrieve it.

Table 4 shows the share of individuals who completed at least one passive and at least one active task, by group of respondents. This makes it possible to assess whether vulnerable individuals are more reluctant to perform one type of task than another. In general, there is a lower propensity for active tasks as a result of a higher burden, regardless of the category to which respondents belong. However, the participation is consistently lower within the vulnerable subgroup, with a more pronounced gap observed for passive tasks (-14.6 percentage points) than for active tasks (-6.8).

Table 4 – *Participation in passive and active tasks – All and digitally vulnerable respondents (% values).*

Task type	All respondents	Vulnerable respondents
≥ 1 passive	49.7	35.1
≥ 1 active	16.0	9.2

Source: Survey on NWM – year 2024

An important insight into digital engagement emerges when comparing respondents' *hypothetical willingness* to perform smart tasks with their *actual behaviour* during the online completion.

Table 5 displays the row percentages of individuals who effectively completed a task out of those who responded to a preliminary question on their hypothetical willingness, by task and group of respondents.

With reference to the general population, the discrepancy between intention and action is substantial across all tasks, with the smallest observed difference[6] being for geolocation sharing (37.5 percentage points) and the largest for meter reading (85 percentage points).

The discrepancy is wider for the respondents who are digitally vulnerable, with differences ranging from 60.6% for the sharing of step counts to 94.8% for the uploading of receipts. In comparison with the general population, a higher proportion of individuals are unable to perform the tasks, whether due to a lack of tools or skills. This suggests that the process of consenting to and completing a smart task poses a

[6] The difference is calculated as 100.0 minus the value at *Hypothetical willingness* = "Yes" and *Real performance* = "Shares" in Table 5.

distinct challenge that many vulnerable individuals are unable to overcome, even when they have previously expressed their intent.

Table 5 – *Real performance against hypothetical willingness to perform the smart tasks – All and digitally vulnerable respondents (row percentages).*

Hypothetical willingness	Real performance					
	All respondents			Vulnerable respondents		
	Shares	*Is not able to*	*Does not share*	*Shares*	*Is not able to*	*Does not share*
Share geolocation						
Yes	62.5	23.2	14.3	35.4	54.8	9.8
Maybe	38.7	35.5	25.8	54.4	26.7	19.0
No	16.9	62.9	20.2	36.2	31.7	32.1
Does not know	31.8	45.5	22.7	18.9	53.1	28.0
Share step count						
Yes	47.3	42.2	10.5	39.4	52.8	7.8
Maybe	42.3	54.6	3.1	18.2	76.5	5.3
No	19.5	67.9	12.6	8.4	77.7	13.9
Does not know	20.9	67.4	11.6	11.4	86.1	2.5
Share photos of shopping receipts						
Yes	18.1	65.1	16.9	5.2	75.2	19.6
Maybe	17.5	71.4	11.0	10.5	52.6	36.9
No	7.3	56.3	36.5	6.7	43.2	50.1
Does not know	9.4	56.6	34.0	6.8	37.0	56.1
Share photos of energy meters						
Yes	15.0	27.6	57.4	14.8	15.6	69.6
Maybe	4.7	26.9	68.4	0.0	21.0	79.0
No	4.5	15.8	79.8	1.7	9.0	89.3
Does not know	2.1	25.0	72.9	0.0	11.7	88.3

Source: Survey on NWM – year 2024

The combination of motivational readiness and structural constraints (*e.g.*, unfamiliarity with device functions, non-possession of instruments, distrust) appears to generate a wider gap between stated and real participation.

The observed discrepancies between hypothetical and actual performance support the well-known distinction in behavioural research (Ajzen, 1991) between attitude and action. While attitudinal openness toward smart surveys may be relatively widespread, actual engagement might depend on factors that include: technical readiness, task comprehension, privacy and trust concerns, cognitive load and task fatigue.

5. Qualitative data analysis of open-ended responses on the reasons for not performing the smart tasks

The NMW-S questionnaire included four open-ended questions aimed at exploring the reasons why respondents do not perform specific smart tasks. The four open-ended responses, one for each smart task, are analysed using an open coding process (Strauss and Corbin, 1998), a method that allows for the identification and development of concepts and their dimensions directly from the data.

According to this bottom-up approach, codes are generated inductively and grouped into categories based on shared thematic elements. These categories are further aggregated into higher-order or main categories. The initial coding schemes, which include both main categories and emergent subcategories, are thoroughly reviewed in order to ensure a consistent interpretation of the responses' underlying meaning. The final coding schemes are then applied to the full set of responses, providing an evidence-based description of the findings.

As stated in the literature, open-ended questions have much higher non-response rates than other types of survey items (Millar and Dillman, 2012). Table 6 provides the percentage of non-response to the four NWM-S open-ended questions for all respondents and the vulnerable subgroup. As shown in the table, the rates by task are high and quite similar for both groups.

Table 6 – *Non-response rates to open-ended questions of NWM-S – All and digitally vulnerable respondents (% values).*

Smart task	All respondents	Vulnerable respondents
Share geolocation	53	57
Share steps count	98	99
Share photos of shopping receipts	68	66
Share photos of energy meters	48	48

Source: Survey on NWM – year 2024

Since the step count task has a high non-response rate in both groups, only the results related to the other tasks, which have a lower non-response rate, are examined in detail below. The first open-ended question was addressed to respondents who did not share their geolocation. The results indicate that the main reason for not performing this task is a lack of trust and information. This is particularly due to privacy concerns and a lack of clarity about the usefulness of the task. Moreover, these motivations are similarly prevalent in both disadvantaged and general groups (Table 7). Besides, approximately 1% of vulnerable respondents reported being unable to complete the smart task due to technical limitations – such as the lack of a suitable device – or personal difficulties that prevented them from proceeding.

Results related to the reasons for not sharing photos of shopping receipts and energy meters are presented in Tables 8 and 9, respectively.

Table 7 – *Reasons for not sharing geolocation – All and digitally vulnerable respondents (% values).*

Main category	Category	All respondents	Vulnerable respondents
Lack of trust and information	Privacy issues	36	32
Lack of trust and information	Unclear relevance	2	2
Lack of trust and information	Unclear purpose	2	1
Technical issues	No equipment	0	1
Personal issues	No knowledge	0	1
Miscellaneous	Other	6	6

Source: Survey on NWM – year 2024

Table 8 – *Reasons for not taking and sharing photos of shopping receipts – All and digitally vulnerable respondents (% values).*

Main category	Category	All respondents	Vulnerable respondents
Lack of trust and information	Privacy issues	19	19
Lack of trust and information	Unclear purpose	1	2
Lack of trust and information	Unclear relevance	2	1
Miscellaneous	Other	10	12

Source: Survey on NWM – year 2024

Table 9 – *Reasons for not taking and sharing photos of energy meters – All and digitally vulnerable respondents (% values).*

Main category	Category	All respondents	Vulnerable respondents
Lack of trust and information	Privacy issues	28	26
Lack of trust and information	Unclear purpose	2	3
Lack of trust and information	Unclear relevance	3	1
Technical issues	No access	5	5
Technical issues	No equipment	0	1
Personal issues	Too much effort	4	4
Personal issues	No time	2	3
Miscellaneous	Other	6	8
Miscellaneous	Irrelevant	1	1

Source: Survey on NWM – year 2024

As shown in Table 8, the reasons for not sharing a photo of the shopping receipt are primarily related to the main category “Lack of trust and information”, indicating a need for clearer communication on data privacy, purpose, and relevance. In contrast, the reasons for not sharing a photo of the energy meter are more diverse (Table 9). While a lack of trust and information is common to both categories of

respondents, technical issues and personal difficulties – often related to uncertainty about the data collection procedure – emerge specifically in the vulnerable subgroup.

6. Conclusions

This study provides insights for designing equitable and effective data collection strategies, as well as for mitigating the non-response bias introduced by the smart surveys in the Italian context.

The logistic model indicates that the respondents who have access to a wider range of devices and to greater financial resources within their households are more likely to engage in smart tasks, especially if they are not concerned that their data might be stolen or misused. It is worth noting, however, that socio-demographic characteristics affecting the online participation – such as educational level and age – turn out to be weak predictors of the smart tasks' performance.

Descriptive analyses highlight a considerable discrepancy between the stated willingness to engage in smart actions and the actual performance of such actions, particularly among the respondents who are digitally disadvantaged. Considering that the digitally vulnerable subgroup represents 31% of the total population, it is crucial to take the digital inequality into account when designing smart surveys in order to improve the participation of individuals and ensure the representativeness of the final data.

Open-ended responses reveal that a lack of trust and information, along with technical and personal issues, are the main reasons why respondents do not complete smart tasks. As it is challenging to act on the structural factors contributing to the digital divide, the most viable short-term strategy to enhance the implementation of smart surveys appears to be the fostering of trust in digital systems. The first step in achieving this objective is to address privacy and security concerns through a clear communication of the measures in place to protect personal information. Other key elements include transparency about data usage and the implementation of robust security protocols.

Acknowledgements

The authors are grateful to Sandra Cipparrone, Viviana De Chirico, Federico De Cicco, Eleonora Deiana and Linda Porciani for their valuable contribution to the implementation of the NWM survey.

References

AJZEN I. 1991. The theory of planned behavior, *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, pp. 179-211.

KEUSCH F., STRUMINSKAYA B., LUGTIG P., HOHNE J.K. 2023. Smart Surveys for the Digital Age: Promise and Pitfalls, *Journal of Official Statistics*, Vol. 39, No. 1, pp. 23–45.

HOSMER JR. D.W., LEMESHOW S., STURDIVANT R.X. 2013. *Applied Logistic Regression*. 3rd edition. Hoboken, NJ, U.S.: Wiley, Series in Probability and Statistics.

LUNARDELLI I., VAN DEN HEUVEL J., SCHOUTEN B., NUCCITELLI A., D'AMEN B., LORÈ B., PEREZ M., ZGONEC M. 2024. Smart Perceptions Survey. *Eurostat project Smart Survey Implementation Deliverable 1.2*, https://cros.ec.europa.eu/book-page/coordination-and-integration-smart-baseline-stage-12

MILLAR M.M., DILLMAN D. 2012. Do Mail and Internet Surveys Produce Different Item Nonresponse Rates? An Experiment Using Random Mode Assignment, *Survey Practice*, Vol. 5, No. 2, https://doi.org/10.29115/SP-2012-0011

MOSSBARGER K., TOLBERT C.J., STANSBURY M. 2003. *Digital Citizenship: The Internet, Society, and Participation*. Cambridge, MA: MIT Press.

SAS INSTITUTE INC. 2015. SAS/STAT® 14.1 User's Guide. Cary, NC: SAS Institute Inc.

STRAUSS A.L., CORBIN J.M. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks: Sage Publications.

STRUMINSKAYA B., LUGTIG P., KEUSCH F., HOHNE J.K. 2020. Augmenting Surveys with Data from Sensors and Apps: Opportunities and Challenges, *Social Science Computer Review*, Vol. 40, No. 2, pp. 220–238.

VAN DEURSEN A.J.A.M., VAN DIJK J.A.G.M. 2014. The digital divide shifts to differences in usage, *New Media & Society*, Vol. 16, No. 3, pp. 507–526.

Monica PEREZ, Istat, perez@istat.it
Barbara D'AMEN, Istat, barbara.damen@istat.it
Barbara Maria Rosa LORÈ, Istat, lore@istat.it
Alessandra NUCCITELLI, Istat, nuccitel@istat.it