

Data Quality in Social Survey Research

D

Marco Palmieri

 <https://orcid.org/0000-0002-9081-0052>

La Sapienza University of Rome, Italy

Rosario Aprile

 <https://orcid.org/0000-0003-1259-3272>

La Sapienza University of Rome, Italy

INTRODUCTION¹

The social survey research community has always felt the need to identify and share criteria for assessing the quality of collected and analysed data. Data quality is a fundamental aspect of social survey research that comes from the diffusion and usage of large-scale surveys planned without adequate methodological design. In the long tradition of reflections on the data quality in surveys, it seems that two antithetical positions prevail: according to the behaviourist paradigm, data quality is the absence of distortions in the measurement process (Groves, 1991); otherwise, the pragmatic position conceptualises data quality as the satisfaction of the logical and methodological conditions necessary for the achievement of the cognitive research's objectives (Mauceri, 2003).

The first position originated in psychometrics, and social survey methodology imported it. The nineteenth century saw the spread of what Von Hayek (1989) called the scientist attitude, according to which the human sciences can develop only by following the guidelines, the method and the procedures of the physical sciences. The gradual devaluation of the philosophical differences between the physical and human sciences has transferred to the social sciences, mainly quantitative social science, giving rise to the so-called survey methodology. This approach is based on measurement theory that conceives data quality as the absence of error in the measurement process: the researcher investigates the sources of error that can move the measured values away from the true values (Groves, 2004).

Other scholars criticise this scientific attitude and claim the social sciences' ontological, epistemological, and methodological autonomy from natural and physical sciences (Marradi, 2016). Rejecting the idea that only one model of science exists (that one of hard sciences) and accusing social scientists of denying their philosophical origins - because the models of the physical sciences attract them - these scholars have given rise to a pragmatic view on data quality debate. The pragmatic approach looks at the actual conditions of research and the cognitive goals that move it. "There is data quality when the researcher succeeds in fulfilling the logical and methodological conditions necessary to achieve the cognitive objectives of the research" (Mauceri, 2003, p. 41).

Analysing these two paradigms is necessary to understand the complex data quality topic.

DOI: 10.4018/978-1-6684-7366-5.ch018

This article, published as an Open Access article in the gold Open Access encyclopedia, Encyclopedia of Information Science and Technology, Sixth Edition, is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

BACKGROUND

The attention to data quality measurement originated in psychometrics, whose roots are in the behaviourism paradigm, born in the early 1900s when the focus of the academic community of psychologists shifted from introspective methods to external environments in order to understand and explain human behaviour. Behaviourism has transmitted its appeal to social survey methodology by attempting to pursue “the comparability of responses” through the invariance of the stimuli and the standardisation of the interview situation (Fideli & Marradi, 1996, p. 74). The mechanical stimulus-response model represents the survey interview as follows: the stimulus (the questionnaire’s question) is characterised by a high degree of structuring, and the respondents have to answer the same questions in the form and order designed by the researcher (Gobo, 1997). Fowler and Mangione summarise the core of the behaviourist manifesto in survey research: “the key element of measurement is standardisation. Standardisation aims to expose all respondents to the same question and response alternatives. In this way, differences can be interpreted as actual differences among respondents” (1990, p. 14). The highly standardised interview theorised by behaviourists requires establishing strict rules, especially regarding the interviewer’s role, which is conceptualised as one of the leading causes of measurement error in the survey. In this approach, the perfect behaviourist interviewer has high skills to follow standardisation rules, administering stimuli designed by others within the cage imposed by standardisation in order not to introduce bias and errors during the interview (Biemer et al., 1991).

According to the measurement theory, the act of measurement in a survey produces an empirical observation, which incorporates an error from the true value. The measurement theory intends to study and tackle the errors from the true values observations. Several authors, such as Deming (1944) and Kish (1965), have contributed to conceptualising the social survey error. Andersen et al. (1979) theorised the total survey error, identifying the effects of the different error sources. Then Biemer and Lyberg (2003) introduced the conceptual specification error (the inadequate and poor specification of object investigation) and the processing error (i.e. data cleaning, coding, data entry, and statistical weighting). For Groves (2004), survey methodology was born to study the error sources that, either randomly or systematically, affect the quality of all survey phases. The scholar has differentiated the systematic error between the observation error (the distance between a subject’s answer and the subject’s true value on the measured property) and the non-observation error caused by statistical estimation error (the distance between the sample value and the population value).

The measurement theory in survey research comes from the positivists’ and neopositivists’ epistemological framework: the task of science is to pursue the truth (objective knowledge), and the scientist’s task is to discover the reality without introducing any element of personal interpretation. This idea of science refers to objectivist gnoseology: the conceptual structures with which the researcher represents the object under investigation are not free creations of the human intellect but are provided naturally by the object itself through the sensations emitted to the perceiving agent. “The etymology of the term ‘data’ emphasises the passive attitude of the researcher” (Bruschi, 1990, p. 224) because the data would be coercively imposed by an objective reality external to the researcher that can be grasped only through scientific activity (Agnoli, 1992, p. 144).

An alternative theoretical approach to the quality of data in social survey research, labelled by Mauceri (2003) as pragmatic, rejects the existence of true data - and the measurement theory - and defines data as the outcome of a process, “the product of operations of selection, collection and formalization” (Statera, 1994, p. 125) that the researcher deploys to achieve the cognitive research goals. Data is the lowest unit of information that, after being collected and formalised, is coded into a matrix and subjected to further

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/data-quality-in-social-survey-research/322092

Related Content

Affordances and Constraints of Automation and Augmentation: Lessons Learned From Development of a Human-AI Collaboration Business Simulation Platform

Qingyu Liang, Juanqiong Gou, Zhe Wang and Marina Dabi (2024). *Journal of Global Information Management* (pp. 1-27).

www.irma-international.org/article/affordances-and-constraints-of-automation-and-augmentation/357260

Developing Country Perspectives on Software: Intellectual Property and Open Source. A Case Study of Microsoft and Linux in China

Xiaobai Shen (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1369-1389).

www.irma-international.org/chapter/developing-country-perspectives-software/19045

Capacity Development Initiatives for Marginal Communities: A Few Case Studies

M. Aminul Islam, Elena Murelli, Frederick Noronha and Hakikur Rahman (2009). *Selected Readings on Global Information Technology: Contemporary Applications* (pp. 110-133).

www.irma-international.org/chapter/capacity-development-initiatives-marginal-communities/28608

Expense and Exposure: Virtual Global Education for Tomorrow

Bobbe Cummins Colburn and Julie Nolin (2012). *Disruptive Technologies, Innovation and Global Redesign: Emerging Implications* (pp. 395-406).

www.irma-international.org/chapter/expense-exposure-virtual-global-education/63842

Developing a Generic Framework for E-Government

Gerald Grant and Derek Chau (2006). *Advanced Topics in Global Information Management, Volume 5* (pp. 72-101).

www.irma-international.org/chapter/developing-generic-framework-government/4562