ОБЩЕСТВЕНИ КОМУНИКАЦИИ И ИНФОРМАЦИОННИ НАУКИ
PUBLIC COMMUNICATIONS AND INFORMATION SCIENCES

MEASUREMENT ERRORS IN QUESTIONNAIRES: AN ANALYSIS OF QUESTION FORMAT-SPECIFIC ERRORS

Daniel Schilling
University of Library Studies and Information Technologies

Abstract: *This paper examines the measurement errors in questionnaires, focusing on question format-specific errors that can affect validity. Through a comprehensive literature review, 13 types of errors have been identified, focusing in particular on multiple-choice and free-answer questions. The results shows that multiple-choice questions, among others, are subject to "test wiseness" and "irrelevant difficulty", while free-answer questions result in a higher rate of non-response. It covers different types of errors, including dead-end questions, double-barrelled questions and loaded questions. The investigation also looks into the strategic use of certain question types to elicit more valid responses. This research shows how question-related errors affect questionnaire trustworthiness. It provides a basis for improving questionnaire design, valuable especially for non-expert practitioners. This paper aims to enhance the validity and reliability of information gathered via this widely used tool by highlighting the importance of careful questionnaire design.*
Keywords: *Free-answer questions, multiple-choice questions, question format errors, question-specific errors, questionnaire validity*

INTRODUCTION

Questionnaires are one of the most important means of gathering information, used for a variety of reasons in private as well as in a professional context. Beside private individuals especially companies, public institutions and even supervisory authorities use this tool to obtain and analyse needed information. Its widespread use is due to its supposedly simple creation, design and use. However, only in few cases are the inherent negative aspects sufficiently taken into account. Scientifically, questionnaires are susceptible to a wide variety of errors and may only be limited suitable for obtaining valid answers. The measurement errors considered in this paper are embedded in a variety of possible errors in the use of questionnaires. Basically, they can be divided into two categories: respondent-specific aspects, and thus in particular cognitive-psychological sources of error, and methodological errors. Cognitive or psychological biases relate, for example, to false memories, inadmissible contextual conclusions or influencing the respondent through the sequence of questions or the answer format. Methodological errors, which include the measurement error considered in more detail, include aspects such as sampling and scoping errors, relevant aspects that where not questioned, errors regarding coverage or non-response and, in particular, information processing errors. The measurement errors examined in this paper pertain specifically to question format errors. Fig. 1 provides an overview of question format-specific errors[1].

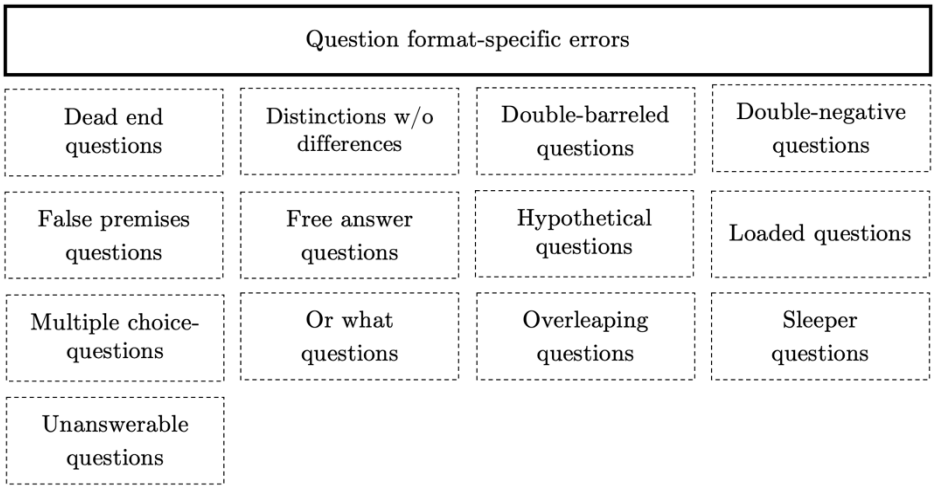

Fig. 1. Overview of question format-specific errors in alphabetical order (figure by the author, based on Payne 2014)

After introducing the methodology, this paper discusses selected question format-specific errors to consider for obtaining accurate and meaningful information from questionnaires.

RESEARCH METHODOLOGY

Various academic fields provide insights into error sources in questionnaires. This paper examines current research in depth. The assessment relies on a wide variety of sources, such as published books, scholarly journals, research papers, and web publications across different disciplines.

RESULTS

Multiple-choice questions are widely utilised in questionnaires, and are for instance the most common method employed in written examinations (Natekar, DeSouza and Karapurkar 2016, p. 558). The frequency with which multiple-choice questions are used is accompanied by the prevalence of construction errors that jeopardise their validity (O'Neill et al. 2019, p. 51). As with any form of closed question, response options that are not listed are rarely considered by respondents in multiple-choice questions, instead, respondents tend to select options that they might not have chosen or might have forgotten otherwise, just because they are stated (Sheatsley 1983, pp. 206–208). Furthermore, specific biases exist in the use of multiple-choice questions. When multiple-choice questions are employed to assess knowledge, two potential sources of error are identified: test wiseness and irrelevant difficulty (Case and Swanson 2002, p. 19). Test wiseness enables the respondent to select the correct or desired answer, without employing the requisite knowledge. This might occur, for instance, if distractors do not follow grammatically from the stem, or if terms such as "always" or "never" are used in the options (Towns 2014, p. 1428). Additionally, it is worth noting that in multiple-choice questions, the correct answer may be chosen purely by chance (Fowler 1995, p. 69). Further pertinent is the construction flaw irrelevant difficulty. This pertains to the poor design of multiple-choice questions in questionnaires. Validity can be compromised, particularly through the use of lengthy, complex, or inadequately distinct answer options, potentially incorporating vague terms, inconsistent presentation of numeric data, or simply complicated wording (Case and Swanson 2002, pp. 22–25). The case where numeric data are not stated consistently is illustrated by the example in Fig. 2. Here, five answer options are used, some of which overlap. In such cases, the validity of the questionnaire is likely to be compromised, and they should therefore be prevented through careful editing and revision. It should be noted that drafting a good multiple-choice question can take an hour or more, even for professional item writers (Albanese and Gjerde 1987, p. 280). Therefore, careful preparation of questionnaires is likely to require several working days. In light of this, there may be a temptation to use free-answer questions instead of multiple-choice questions. However, Bradburn (1983, p. 300) concludes that free-answer questions are not superior in eliciting responses of greater validity.

Tick	How long is the guaranteed response time for level 1 security incidents?
☐	Under 1 h
☐	30 – 60 min
☐	¼ working day
☐	1-3 h
☐	4 h to 300 min

Fig. 2. Example for numeric data that is unclear and not stated consistently in multiple-choice questions (figure by the author)

The benefits associated with **free-answer question** formats, including diminished biases due to the absence of predefined response options and the capacity to capture nuanced and complex security measures (Connor-Desai and Reimers 2019, p. 1427), are offset by certain drawbacks that may undermine the validity of such questions. The necessity of coding responses introduces the potential for errors at this stage. Moreover, free-answer questions occupy a unique position with respect to non-response error, as their implementation is often associated with a higher rate of abandonment (Connor-Desai and Reimers 2019, p. 1438). Even if there is a compulsion to answer, abandonment is to be expected, for example in the form of answers being given at random without reflection on the content. One reason for this is the greater time required to respond (Vicente and Reis 2010, pp. 260–261). Andrews (2005, p. 3760) reported non-response rates for free-answer questions of up to 76%, while the average rate across all questions was merely about 2%. Respondents who choose to answer free-answer questions, particularly when these are not mandatory, tend to be disproportionately those who are dissatisfied and provide negative feedback (Macey 1996, p. 229), but the majority of respondents do not answers free-answer questions (Griffith 1999, pp. 1002–1004). This phenomenon, known as positive-negative asymmetry, imply that responses to free-answer questions are disproportionately negative.

Further to be considered are **dead-end questions**, defined as queries that cannot be answered meaningfully, as they do not provide an opportunity for a truthful and accurate response. They often contain problematic keywords such as “never”, “all” or “every” (Payne 2014, pp. 192–193). A pertinent example in the professional field might be questions such as “Are all possible measures implemented?” or “Are all legal regulations adhered to as comprehensively as possible?”. Although the issues inherent in these examples may appear evident, such questions are commonly encountered in practice and even in legal directives, see e.g. Art. 28 (5) EU Digital Operation Resilience Act from 2023.

This analysis pertains to the difficulties that can arise when questions cannot be meaningfully distinguished in terms of their content. For example, the questions “When was this measure implemented?” and “Since when has this measure been in place?” may not offer a substantive difference in terms of their content. Such indistinguishability may lead respondents to question the validity of the questionnaire, potentially resulting in less truthful or precise responses, and an inclination to infuse personal interpretations and contextual information (Schober and Clark 1992, pp. 27–28). An indication that respondents may be struggling to differentiate between questions, or that the questions are contextually redundant, can be observed when respondents annotate their responses with comments such as “Answer see question […]”. The intentional use of similar questions as sleeper questions to detect inattentive respondents will be addressed later.

When two distinct issues are addressed within a single query, it is termed a **double-barrelled question** (Sheatsley 1983, p. 216). This presents a challenge for respondents, particularly when the format is not free-answer. Consider the question: “Does the organisation ensure that all data is encrypted and backed up regularly?” In this case, the question conflates two issues, making it difficult for the respondent to provide a precise answer if they agree with one aspect but not the other. It is noteworthy that this example of a double-barrelled question is additionally a dead-end question, as it queries “all data”.

Double-negative errors arise when a question is framed with two negative terms, potentially leading to increased cognitive effort required for comprehension and subsequent misinterpretation by the respondent. Over a century ago, Wembridge and Means (1918) investigated ambiguities in voting due to the use of negatives and found that positively framed questions were more easily understood, whereas their negative counterparts were deemed highly confusing. Similarly, Blankenship (1943, p. 59) concludes unsurprisingly, that phrasings expressed positively are more likely to be comprehended correctly.

A **false premise error** occurs when a question is based on an assumption that is not entirely or only partially true, potentially leading to biased responses and incorrect information from the respondent. An example might be a question concerning the frequency of salary increase, which presupposes that these increases are happening.

When using **hypothetical questions**, questions concerning future scenarios rather than actual past events or current practices, it is important to recognise that responses to these questions, while appearing reasonable, often do not align with the actions or behaviours that would occur if the proposed scenario were to materialise and ultimately may not reflect the truth (Payne 2014, p. 198; Kaderabek and Sinibaldi 2022). This includes "what-if" questions and questions regarding the intention to act, which Sheatsley (1983, p. 217) has even described as "useless and worthless as predictors". Although it may sometimes be necessary to pose hypothetical questions, it should be noted that respondents are generally poor at predicting the future, including their own future behaviour (Payne 2014, p. 198). This difficulty arises due to changing circumstances and the many situational variables that intervene (Sheatsley 1983, p. 218).

Loaded questions hold a special position due to their inherent nature of suggesting a particular response, thereby typically making respondents' effort to disagree greater than to agree. However, the creator of a questionnaire can leverage this phenomenon. Kinsey et al. (2023, pp. 53–54) highlight that loaded questions can decrease the burden of admitting negative aspects. Thus, a question such as "Many organisations may not implement all required security measures. Which requirements have not yet been implemented in your organisation?" might be better suited than unloaded questions to identify missing security measures. By phrasing the question this way, the respondent is indirectly informed that missing measures are considered normal and that it is acceptable to report them (Sheatsley 1983, p. 215).

The potential use of **sleeper questions** aims to identify inattentive respondents by repeating similar questions multiple times to assess whether the respondent provides consistent answers. Further possible is the providing of answer options that are non-existent. A classic example shows that approximately 70% of respondents provided answers related to a fictitious act, thus being identified as sleepers (Gill 1947 as cited in Westle 2024, p. 238). However, later studies have reported a lower percentage, with Schuman and Presser (1980, p. 1214) noting figures between 25% and 30%. It is important to note that such attentiveness checks carry the risk that respondents who recognise the test may either terminate the questionnaire or alter their responses.

CONCLUSION

A large number of errors related to question format were identified, which are also interrelated and must be taken into account when designing questionnaires. For many of the biases discussed, it can be noted that experts in question design are less likely to make them. However, the creators of business-related questionnaires are generally not experts in question formulation, and thus seemingly obvious and simple mistakes still have relevance and need to be considered. Siddiqui (2024, p. 36) found that 38% of multiple-choice questions had errors, and 12% of the questions even had multiple errors. It was further shown that selected phenomena can be utilised and how this can be used to obtain necessary information. This paper provides an introduction to research on how these question-related errors influence the trustworthiness of questionnaires and forms the basis for further considerations regarding detailed design specifications for questionnaires.

NOTES

1. Due to the large number of sources of error, it is not possible to give equal weight to all possible negative influencing factors. Given their significant relevance in professional context, multiple-choice questions and free-answer questions are accorded priority and are examined in greater detail. Or what-questions and unanswerable questions are not discussed in this paper.

REFERENCES

ALBANESE, M. A., C. L. GJERDE, 1987. Evaluation. In: H. L. Van HOOZER (ed.). *The Teaching process: theory and practice in nursing.* Norwalk, Conn: Appleton-Century-Crofts, pp. 269–308.

ANDREWS, M., 2005. Who Is Being Heard? Response Bias in Open-ended Responses in a Large Government Employee Survey. In: *AAPOR – ASA Section on Survey Research Methods.* Alexandria, pp. 3760–3766.

BLANKENSHIP, A. B., 1943. *Consumer and Opinion Research.* New York, London: Harper & Brothers.
BRADBURN, N. M., 1983. Response effects. In: ROSSI, P. H. and J. D. WRIGHT (eds). *Handbook of survey research.* New York: Academic Press, pp. 289–328.
CASE, S., D. SWANSON, 2002. *Constructing Written Test Questions For the Basic and Clinical Sciences.* Philadelphia: National Board of Medical Examiners.
CONNOR-DESAI, S., S. REIMERS, 2019. Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. *Behavior Research Methods,* 51(3), pp. 1426–1440. doi: 10.3758/s13428-018-1066-z.
FOWLER, F. J. Jr., 1995. *Improving survey questions: design and evaluation.* Thousand Oaks: Sage Publications, Inc.
GRIFFITH, L., 1999. Comparison of Open and Closed Questionnaire Formats in Obtaining Demographic Information From Canadian General Internists. *Journal of Clinical Epidemiology,* 52(10), pp. 997–1005. doi: 10.1016/S0895-4356(99)00106-7.
KADERABEK, A., J. SINIBALDI, 2022. Assessing Measurement Error in Hypothetical Questions. *Survey Practice* [Preprint]. doi: 10.29115/SP-2022-0010.
KINSEY, A. C., W. B. POMEROY, C. E. MARTIN, J. A. ALLEN, 2023. *Sexual behavior in the human male.* Bloomington: Indiana University.
MACEY, W. H., 1996. Dealing with the data: Collecting, processing, and analysis. In: A. I. KRAUT (ed.). *Organizational surveys.* London: Jossey-Bass, pp. 204–232.
NATEKAR, P. E., F. DeSOUZA, N. KARAPURKAR, 2016. Errors in MCQ banks: A need to improvise the quality of multiple choice questions. *Journal of the Anatomical Society of India,* 65, pp. 58–61. doi: 10.1016/j.jasi.2016.05.005.
O'NEILL, L. D., S. M. MORTENS, C. NØRGÅRD, A. L. ØVREHUS, U. G. FRISS, 2019. Screening for technical flaws in multiple-choice items. A generalizability study. *Dansk Universitetspædagogisk Tidsskrift,* 14(26), pp. 51–65. doi: 10.7146/dut.v14i26.106496.
PAYNE, S. L. B., 2014. *The Art of Asking Questions.* Princeton: Princeton University.
SCHOBER, M. F., H. H. CLARK, 1992. Asking questions and influencing answers. In: *Questions about questions: Inquiries into the cognitive bases of surveys.* New York: Russell Sage, pp. 15–48.
SCHUMAN, H., S. PRESSER, 1980. Public Opinion and Public Ignorance: The Fine Line Between Attitudes and Nonattitudes. *American Journal of Sociology,* 85(5), pp. 1214–1225.
SHEATSLEY, P. B., 1983. Questionnaire Construction and Item Writing. In: P. H. ROSSI, D. JAMES and A. B. ANDERSON (eds). *Handbook of survey research.* 5th edn. San Diego: Academic Press, pp. 195–230.
SIDDIQUI, Z. S., 2024. Errors in the construction of multi-choice questions: an analysis. *Pakistan Journal of Medicine and Dentistry,* 7(3), p. 4.
TOWNS, M. H., 2014. Guide To Developing High-Quality, Reliable, and Valid Multiple-Choice Assessments. *Journal of Chemical Education,* 91(9), pp. 1426–1431. doi: 10.1021/ed500076x.
VICENTE, P., E. REIS, 2010. Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. *Social Science Computer Review,* 28(2), pp. 251–267. doi: 10.1177/0894439309340751.
WEMBRIDGE, E. R., E. R. MEANS, 1918. Obscurities in voting upon measures due to double-negative. *Journal of Applied Psychology,* pp. 156–163.
WESTLE, B., 2024. Politisches Wissen und Pseudowissen in Online-Erhebungen: Raten und Recherchieren. In: B. WESTLE and M. TAUSENDPFUND (eds). *Politisches Wissen: Korrekte Kenntnisse, Fehlvorstellungen und Ignoranz.* Wiesbaden: Springer Fachmedien, pp. 227–262. doi: 10.1007/978-3-658-42979-9_8.

ГРЕШКИ ПРИ ИЗМЕРВАНЕТО ВЪВ ВЪПРОСНИЦИТЕ: АНАЛИЗ НА ГРЕШКИТЕ, СПЕЦИФИЧНИ ЗА ФОРМАТА НА ВЪПРОСА

Резюме: *В тази статия се разглеждат грешките при измерването във въпросниците, като се обръща внимание на грешките във формата на въпроса, които могат да повлияят на валидността. Чрез обстоен литературен преглед са идентифицирани 13 вида грешки, като се обръща специално внимание на въпросите с избираем отговор и въпросите със свободен отговор. Резултатите показват, че въпросите с избор между няколко отговора, наред с други, са обект на „мъдрост на теста", докато въпросите със свободен отговор водят до по-висок процент на неотговаряне. Обхванати са различни видове грешки, включително въпроси тип „задънена улица", въпроси с двоен отговор и натоварени въпроси. Разследването разглежда и стратегическото използване на определени видове въпроси за получаване на по-валидни отговори. Това изследване показва как грешките, свързани с въпросите, влияят върху надеждността на въпросника. То осигурява основа за подобряване на разработването на въпросници, което е ценно за практикуващите специалисти, които не са експерти. Целта на настоящата статия е да се повиши валидността и надеждността на информацията, събрана чрез този широко използван инструмент, като се подчертае значението*

на внимателното разработване на въпросника.
Ключови думи: *въпроси със свободен отговор, въпроси с избор между няколко отговора, грешки във формата на въпроса, специфични грешки във въпроса, валидност на въпросника*

Даниел Шилинг, докторант
Университет по библиотекознание и информационни технологии
E-mail: schilling@mail.bg