REVIEWS

Biased measure, strong evidence?–An overview of common risk of bias in measurement

Geneviève Laporte*[1,2], Marilyn Aita[1,2,3]

[1] *Faculty of Nursing, Université de Montréal, Montréal, Québec, Canada*
[2] *CHU Sainte-Justine Research Center, Montréal, Québec, Canada*
[3] *Quebec Network for Nursing Intervention Research, Montréal, Québec, Canada*

Received: December 9, 2020 **Accepted:** January 5, 2021 **Online Published:** January 20, 2021
DOI: 10.5430/jnep.v11n5p40 **URL:** https://doi.org/10.5430/jnep.v11n5p40

ABSTRACT

Although nurses, whether researchers or clinicians, may use measuring instruments in their daily practice, instruments deemed credible sometimes present with several undisclosed biases. These biases can undermine the credibility of the results obtained from their use in research or practice. This article aims to synthesize the most frequent biases of instruments to allow nurse researchers and clinicians to recognize them when exposed to new instruments or undertake an original instrument's development. The types of biases and relevant management strategies are classified into four categories: conceptual, methodological, response and contextual. The strategies recommended by measurement experts address biases introduced in developing, testing, and validating instruments. This article provides an overview of recommended practices for their development and testing. It is expected that this article will contribute to raise awareness of nurse researchers and clinicians towards the possible limitations and biases in using instruments and refine their critical thinking about measurement in their respective fields.

Key Words: Measurement, Bias, Nursing methodology research, Psychometrics, Validity, Reliability

1. CONTEXT AND OBJECTIVES

The rigor and value of a study using quantitative methods are greatly affected by the quality of the instruments used to collect data. For example, questions could be raised about the findings of a large and complex study if the instruments used were of poor psychometric quality. In nursing, both clinical practice and research fields frequently use instruments. Whether these instruments are related to physiological indicators (e.g., blood pressure), diagnosis or screening (e.g., postpartum depression, psychological distress),[1] their quality is imperative to ensure that accurate conclusions are drawn from their measurement. As such, it is crucial to investigate and comprehend the quality of instruments used in nursing research or clinical settings. As a result, this article aims: 1) to identify and synthesize the common risk of biases in the use of instruments and 2) to propose concrete strategies to control these biases.

This article adds onto existing knowledge by reviewing potential risk of bias, classifying them into broad categories and identifying accepted methods to minimize their impact. It also allows for researchers better understanding of the effect of methodological choices upon instrument quality and results as it links these choices to inherent risks.

2. IMPLICATIONS OF BIAS

With the intention of measuring a variable or phenomenon of interest comes the obligation to select an instrument capable

*__Correspondence:__ Geneviève Laporte; Email: genevieve.laporte.1@umontreal.ca; Address: Faculty of Nursing, Université de Montréal, Montréal, Québec, Canada.

of accurately and precisely operationalizing it. In addition, to guide appropriately professional practice in clinical settings, the instrument must be chosen to best suit the context and target population. The selection of an instrument is crucial as it will influence the results obtained by administering these instruments and, consequently, the subsequent clinical practices that will be performed.[2] Ultimately, the selection of one or another instrument could impact the targeted population's health and the quality of care.[1]

3. Identification of sources of bias

Bias is defined as "an influence producing a distortion or error".[3] Several sources of bias are expected to emerge during the development, validation and use of an instrument. For this synthesis, these bias sources have been grouped into four broad categories: conceptual, methodological, response and contextual. Their respective definitions and the strategies described in this manuscript are listed in Table 1.

Table 1. Summary table of risk of bias definitions and associated strategies

	Description	Targeted phase	Source of bias	Risk management strategy
Conceptual biases	Conceptual biases are **systematic biases related to the theoretical and conceptual underpinnings** of the instrument developed.	(D)	• Lack or incompleteness of a theoretical model	• Judicious choice of theoretical foundations • Clear statement of theoretical and operational definitions
Methodological biases	Methodological biases are introduced by **methods used by researchers** to develop and validate the instrument.	(D) (V)	• Incomplete generation of items • Item redundancy • Sampling	• Inductive and deductive process • Content validation • Pre-testing in a similar population • Precise description of the sample • Sampling strategies to ensure representativeness
Response biases	Response bias is introduced when participants **complete the instrument in a biased manner.**	(V) (U)	• Social desirability, faking good • Deviation, faking bad • Acquiescence • Central tendency • Extreme response • Proximity	• Limit number of items • Clear and simple wording • Trend towards anonymity • Adding cognitive interviews to the content validation process • Offer nuanced response options • Randomization of items in the absence of subscales
Contextual biases	Contextual biases are biases **induced by the environment or context** in which the instrument is administered.	(V) (U)	• Professional training • Administration context	• Systematic training of professionals • Adding guidelines or instructions for use • Pre-testing in the target environment • Selection of a suitable time of administration

Note. (D): Instrument development phase; (V): Validation phase; (U): Instrument use phase.

3.1 Conceptual biases

The first source of bias emerges mainly from developing instruments, during which the circumscription and definition of the concept to be operationalized by the instrument take place. Conceptual biases are systematic biases related to the theoretical and conceptual underpinnings of the instrument being developed. These theoretical foundations are mobilized during the development stages of the instruments and will influence the choice of specific dimensions and items. The lack of a theory, the use of incomplete theories, or an unclearly defined concept can have repercussions on the validity of the instrument.[4] Validity refers to the instrument's ability to capture the concept or phenomenon it is intended to measure.[5] For example, if the aim is to measure psychological distress in intensive care, using an instrument that documents depression would result in an incomplete and inaccurate representation of the phenomenon to be captured.[6] Indeed, in this context, depression is one of the many symptoms of psychological distress.[6] Thus, an instrument that estimates only part of the concept to be measured will introduce a bias when it is used. In other words, the scores obtained on the instrument will be influenced by the omission of specific essential dimensions of the concept of interest, which would introduce a conceptual measurement bias.

3.2 Methodological biases

Methodological biases are introduced by the researchers' methods to develop and validate the instrument and are generally predictable. In the stage of the development of the instrument, the generation of items may be the source of several methodological biases. In particular, the sources that support item generation and formulation can potentially introduce systematic bias and threaten the validity of instruments. For example, if a symptom is over-represented in the literature compared to patients' experience, an instrument based solely on the literature will perpetuate this over-representation. Furthermore, pejorative or negative wording or the use of neologisms or acronyms from clinical practice (e.g. “PICC” [Peripherally Inserted Central Catheter]) may influence the responses given by the participants, which would introduce a bias.

Furthermore, the choice of items making up the final version of the instrument is likely to introduce a systematic bias according to the sources used to justify the items selected (e.g. conceptual models, experts, synthesis of the literature). In addition, items may be redundant, implying items that do not contribute substantially to measuring the targeted phenomenon.[7]

Also, the methods used to recruit study participants can be a source of bias since the psychometric qualities reported are specific to the population in which they were tested. Thus, the characteristics of the sample when the psychometric properties of the instruments were established and its comparability to the population targeted by the instrument's use impact the validity of the conclusions drawn from the administration of the instrument.

Thus, methodological choices when developing instruments should be considered with attention when assessing an instrument's psychometric qualities.

3.3 Response bias

Response bias can be defined as bias introduced "when participants respond in a biased manner in response [to an element of the context]"[3] (p.282). Schwarz and Oyserman[8] describe five cognitive steps necessary for the response process, each of which implies the possibility of bias: (I) Understanding the question (II) Remembering the events, beliefs and attitudes in question (III) Estimating their value (IV) Articulating their estimate with the available response options (V) Correcting the answer.[8] Streiner, Norman[7] cite six possible types of response bias: social desirability bias, faking good, deviation bias, faking bad, acquiescence bias, and central tendency bias. To these, Polit and Yang[1] add extreme response bias and proximity bias.

While social desirability bias refers to a person's unconscious tendency to respond in a way that makes him or her look good, faking good bias is introduced by choosing to respond inaccurately to items on the instrument to misrepresent one's condition.[7] Thus, social desirability bias can be introduced at the stage of estimating the value of an attribute, where the overestimation or underestimation of the attribute is more positively perceived socially. By contrast, faking good bias is introduced at the stage of correcting the response, where the respondent deliberately provides a false value in order to gain a benefit. Several explanations have been proposed for these two biases, including the idea that people generally consider themselves superior to the “average person”, which leads them to overestimate their qualities unconsciously or to want others to see them as superior to the average person.[7] One consequence of this bias may be the emergence of tendencies for certain subgroups, for example, where a behavior is considered positive for women but negative for men.[1] Then, in contrast, deviation bias refers to a person's tendency to respond instinctively in a way that deviates from the expected norm, and refers to an individual's choice to present themselves in a bad light in order to gain a benefit.[7] Social desirability and deviation bias can be perceived within the same person at different times, which Streiner, Norman[7] identifies as the "hello-goodbye effect". This effect is where a person presents himself or herself, consciously or unconsciously, in a bad light (deviation bias) to initially impress healthcare professionals and quickly improve to please healthcare professionals in their evaluation (social desirability bias).

The second type of response bias is identical response patterns, where one can find acquiescence bias, the tendency to consistently answer either positively or negatively regardless of the question asked,[7] central tendency bias, the tendency to avoid selecting extreme response choices,[7] and extreme response bias, the tendency to respond at both extremes of a scale even when the respondent does not feel extreme about the item.[1] It should be noted that these biases are mostly introduced when the instrument is particularly long, difficult to complete, or when the participant has little interest, which prompts the participant to make the task easier and quicker to complete.[7] Identical response patterns are problematic because they do not reflect the value as estimated by the respondent, which influences the conclusions drawn from the scores obtained.

The third type of bias is proximity bias, introduced when the participant's response is influenced by their response to a previous item and their willingness to remain consistent.[1] This bias is introduced at the response correction stage during the cognitive response process and leads to an overestimation of the internal consistency of an instrument.[1]

3.4 Contextual biases

Contextual biases are induced by the environment or context in which the instrument is administered. For example, professionals' training prior to the use of an instrument and disparities between the practices of these professionals can introduce certain biases. For example, the guidance provided by a professional during the completion of the instrument could influence the participant's responses.[7] Finally, the context of administration may also affect the response pattern of participants at the time of administration. For example, a context of change or crisis, such as a sudden deterioration in health status or a transition to a new unit, could unconsciously influence the participant's responses.[2] This component must be taken into account when selecting the target period and the method of administration of the instrument.

4. Risk of bias management strategies

Given the potential for bias, researchers and clinicians need to be able to identify and replicate concrete strategies to limit the risk of measurement bias. When properly applied, these strategies can also be part of the arguments considered in the choice of an instrument. Finally, it is expected that researchers and clinicians can implement these strategies in their respective settings and limit the introduction of bias when using them with the target clientele.

4.1 Conceptual biases

The risks of conceptual bias can be controlled by the rational choice of relevant theoretical foundations by the original authors of an instrument.[7] Waltz, Strickland[2] reports on the importance of a theoretical definition of the concept of interest and its significant difference with the operational definition of the concept. The theoretical definition should be based on the literature and anchored in a theory consistent with its use within the authors' own discipline. However, a definition derived from a theory from another discipline must be carefully analyzed to ensure its consistency with the concept's use in the discipline of interest.[2] Furthermore, the operational definition refers more to how the concept of interest can be measured.[3]

By making the theoretical and operational definitions of the concept being measured explicit, the authors of instruments contribute to the transparency of the theoretical foundations, allowing the instrument users to assess them critically. In particular, they can evaluate the coherence of the theoretical and operational definitions chosen with the reality of the context in which they wish to use the instrument. If the definitions do not reflect the concept in the desired context, it is then very likely that the instrument is not valid in the target population and that the results obtained must be questioned.

In the case where the concept of interest is labelled as immature, the process of developing an instrument should be delayed until the concept is mature to avoid proposing instruments that are invalid due to incomplete operationalization.[2] It is also suggested that users of instruments should be attentive to the maturity of the concepts they are trying to evaluate, by retracing its conceptual evolution in the literature, and avoid using instruments where the maturity of the concept is questioned.[2]

4.2 Methodological biases

The potential biases introduced by item generation, selection and formulation can be overcome through a rigorous process incorporating a comprehensive content validation step.[3] The strategies used to generate items should be both inductive and deductive, i.e., the items' content should be grounded in the literature and supported by an exploratory method with the target population, such as interviews or focus groups.[9] Subsequently, it is recommended that an evaluation committee of between eight and twelve experts be formed,[10] including individuals from the target population and experts on the concept of the instrument.[11] By including these experts, the authors of the instrument can benefit from additional advice to evaluate the items generated according to their representativeness of theories and experience, their relevance to the measurement of the concept and the clarity of their wording.[11] In addition, the content validation process can help control bias in the formulation of items. It is also recommended that the instrument be pre-tested to identify problematic items so that they can be modified as early as possible in the process.[9] Thus, a rigorous content validation process can limit the introduction of many systematic biases related to the development stages of the instrument. Therefore, researchers and clinicians assessing the potential of an instrument should pay attention to the content validation process undertaken by the authors of the instrument they are considering.

Second, it was discussed that sampling biases could influence the results obtained with respect to the validity and reliability of the instrument in a given population. Indeed, given the specificity of the properties of an instrument,[7] researchers and clinicians who wish to use an instrument must ensure that the properties are appropriate for the population of interest. Thus, ideally, researchers should implement sampling strategies that optimize the generalization of their conclusions to a sufficiently specific population. Among these, the type of sampling chosen and the investigation of participants' characteristics that may act as a confounding variable in the data analysis are recommended strategies to limit the introduction of bias in relation to the sample.[3] Although probability sampling strategies are rarely possible, sampling should be conducted in a way that is as representative as possible of the population of interest.[1] To this end, multi-site recruitment, stratification, or cluster sampling are strategies that optimize the representativeness of the sample when compared to the population of interest.[1] In addition, it remains crucial for researchers to be transparent in describing the sample in which the instrument was tested to limit the risk of error associated with using the instrument in a distinct population.[3]

4.3 Response bias

Although they can be challenging to eliminate, there are several strategies to limit response bias. These strategies are often easy to implement since they are linked to methodological considerations or item wording.

For example, social desirability and deviation bias can be primarily controlled by making questionnaires anonymous, which is sometimes possible in large surveys but less feasible when administering in person or by telephone.[1] For this type of administration, it is preferable to explain to participants that there are no right or wrong answers, normalize their experience, and encourage honesty by building trust with respondents.[1] Finally, it is also possible to obfuscate the concept that the instrument is intended to measure by giving it a non-specific title.[7] However, this strategy may have ethical implications that should be considered before its use.

Second, specific strategies can limit acquiescence bias, but these are not agreed upon by measurement experts. Indeed, it was once proposed to create instruments where some items were phrased positively and others negatively, but this practice has been discouraged to date.[1] DeVellis[12] recommends that items be formulated in such a way as to standardize their direction. However, if some items are worded negatively, it is recommended that the instrument be pre-tested and that cognitive interviews be undertaken, i.e., participants are asked to explain their response process,[3] to identify the potential for problems.[1] Finally, it was also discussed, not without controversy, that inserting a "surprise" item would identify individuals with a tendency to nod, for example, an item asking respondents to enter a particular response or refrain from responding to that item.[1] However, this strategy does not limit the introduction of bias but instead identifies the respondents exhibiting it, which may still be useful for researchers in subsequent steps. It is also recommended that the number of items be limited and that their clarity and simplicity be optimized to avoid inducing bias of identical responses among participants due to an instrument that is tedious to complete.[7]

Central tendency and extreme response biases are more difficult to control since they usually reflect the respondent's personality trait.[1] However, it is possible to nuance response options ("almost always" rather than "always") to promote the use of the full range of response options.[7] These biases also vary from culture to culture and should be considered when cross-culturally adapting instruments or comparing psychometric properties between two cultural adaptations.[1]

Lastly, proximity bias is influenced by the order in which items are completed. Thus, the randomization of items to compose the instrument would be the most apparent strategy to overcome this risk. However, for instruments with several subscales, it is also possible to space out items from the same subscale to distance items targeting the same construct.[1] This strategy is simple to implement and has, as its only drawback, the complexity of calculating scores for each of the subscales, although these are most rarely of interest.[1]

4.4 Contextual biases

Lastly, since environmental conditions introduce contextual biases, many can be prevented by careful planning of the study and knowledge of the instrument's environment of use.

First, biases related to professionals' training can be limited by systematic training offered to those likely to administer the instrument to the target population. In order to make the process reproducible, it is even recommended that specific guidelines be established regarding administration standards for the instrument to make it as systematic as possible, despite the presence of several professionals.[2] Simple instructions, supported by examples, clarifying the procedure for completing the instrument should be added to guide participants and reduce the response process's variability.[13] Another strategy is to pre-test the process of administering the instrument to ensure that it runs smoothly, especially if professional intervention is required.[9] Pre-testing can also be used to evaluate the administration process in the specific context (environment, physical environment, process, time required) and to minimize differences between professionals.[9] Finally, recruiting practitioners with similar characteristics (level of education, profession, experience) can ensure the consistency of the interventions carried out when administering the instrument.

 ISSN 1925-4040 E-ISSN 1925-4059

Finally, the context of administration is a source of bias that may be difficult to control or modify since the context may evolve within the same environment during the study or may be difficult to alter to allow the instrument's administration. However, the environment should be as neutral as possible to limit exposure to confounding variables. The administration method (in-person, telephone, mail, online) should be consistent with the context and purpose of data collection.[7] The administration method should be chosen based on its advantages and disadvantages (see Streiner, Norman[7] for more information) with consideration of the potential for bias introduced by this method.

5. CONCLUSION

This article provides a synthesis of the different sources of bias that can influence the quality of the available instruments and proposes risk management strategies to minimize their impact on the results obtained. This review thus contributes to the knowledge of researchers and clinicians interested in measurement, making them aware of the possible limits of the instruments they use and sharpening their critical thinking regarding their use in their respective activities.

ACKNOWLEDGEMENTS

The authors wish to thank Pr. Sylvie Le May, Vice-Dean of Research and International Development and Full Professor, Faculty of Nursing, Université de Montréal, and Christian Vincelette, Ph.D. candidate, Université de Sherbrooke, for their contribution to this manuscript.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest.

REFERENCES

[1] Polit DF, Yang F. Measurement and the measurement of change: A primer for the health professions. Philadelphia: Wolters Kluwer; 2016. 350 p.

[2] Waltz CF, Strickland O, Lenz ER. Measurement in nursing and health research. New York, NY: Springer Publishing Company, LLC; 2017.

[3] Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. 10th Edition ed. Philadelphia: Wolters Kluwer Health; 2017.

[4] Rosas SR, Camphausen LC. The use of concept mapping for scale development and validation in evaluation. Eval Program Plann. 2007; 30(2): 125-35. PMid:17689319 https://doi.org/10.1016/j.evalprogplan.2007.01.003

[5] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010; 63(7): 737-45. PMid:20494804 https://doi.org/10.1016/j.jclinepi.2010.02.006

[6] Laporte G, Aita M. Psychological distress as a unique concept: A response to Molgora et al. (2020). J Affect Disord. 2020; 278: 13-4. PMid:32949868 https://doi.org/10.1016/j.jad.2020.09.050

[7] Streiner D, Norman G, Cairney J. Health Measurement Scales A practical guide to their development and use. 5 ed. Oxford, United Kingdom: Oxford University Press; 2015. https://doi.org/10.1093/med/9780199685219.001.0001

[8] Schwarz N, Oyserman D. Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. American Journal of Evaluation. 2001; 22(2): 127-60. https://doi.org/10.1177/109821400102200202

[9] Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Front Public Health. 2018; 6: 149. PMid:29942800 https://doi.org/10.3389/fpubh.2018.00149

[10] Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Research in Social and Administrative Pharmacy. 2019; 15(2): 214-21. PMid:29606610 https://doi.org/10.1016/j.sapharm.2018.03.066

[11] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 2007; 30(4): 459-67. PMid:17654487 https://doi.org/10.1002/nur.20199

[12] DeVellis RF. Scale development : theory and applications. Fourth edition.. ed: Los Angeles: SAGE; 2017.

[13] Morgado FFR, Meireles JFF, Neves CM, et al. Scale development: ten main limitations and recommendations to improve future research practices. Psicologia: Reflexao e Critica. 2017; 30(1): 3. PMid:32025957 https://doi.org/10.1186/s41155-016-0057-1