MEASUREMENT INVARIANCE DEBATE: WHERE IT'S AT AND WHAT IT MEANS FOR TEACHING RESEARCH METHODS

ДИСКУСІЯ ПРО ІНВАРІАНТНІСТЬ ВИМІРЮВАННЯ: СУЧАСНИЙ СТАН ТА ЗНАЧЕННЯ ДЛЯ ВИКЛАДАННЯ ДОСЛІДНИЦЬКИХ МЕТОДІВ

UDC 303.023.032:[008:001.891]
DOI https://doi.org/10.32782/hbts.75.1.3

Maltseva K.S.
Doctor of Sociological Sciences,
Associate Professor,
Head of the Department of Sociology
National University of "Kyiv-Mohyla Academy"

The issue of measurement in cross-cultural research continues to give rise to lively and complex discussions. The theme of validity in cross-cultural comparisons is contingent on evidence of comparability of metrics and data. Specifically, there has been a long-standing debate regarding measurement invariance which centers around the effectiveness (and legitimacy) of meaningfully comparing social, cultural, and psychological constructs across different societies without loss of meaning or its distortion. The issue of measurement invariance itself has generated a large literature. "Measurement invariance" refers to statistical criterion typically considered necessary for using a metric in more than one cultural context. It indicates whether or not an instrument assesses the same construct across different conditions or samples. Although its advocates insist on the necessity of measurement invariance for conducting valid multigroup comparisons, its critics cast doubt on the underlying premises on the grounds of measurement invariance requirements being too stringent, prohibitionist, or misguided. The landscape of measurement invariance debate is inherently complex. This publication systematizes current views on measurement invariance to present a historiographic account of the emergence of different perspectives on measurement invariance debate in social science. By unpacking the complex layers surrounding the measurement invariance issue, as well as its criticisms and relevant recommendations, this review highlights key distinctions in the prevailing perspectives and explores their implications for the current requirements for measurement invariance in empirical research. This debate has important implications for teaching research methods; emerging insights could be used to inform and enrich research design and methods curriculum.
Key words: *measurement, measurement invariance, validity, quantitative methods, culture, cross-cultural research.*

Проблема вимірювання в крос-культурних дослідженнях продовжує сприяти жвавим та комплексним обговоренням. Тема валідності у крос-культурних порівняннях залежить від доказів можливості порівнювати метрики та дані. Зокрема, йдеться про довготривалу дискусію стосовно інваріантності вимірювання, що зосереджується на ефективності (а також самої можливості) робити змістовні порівняння соціальних, культурних та психологічних конструктів між різними суспільствами без утрати смислу чи уникаючи його викривлення. Проблематика інваріантності вимірювання породила значну за обсягом літературу. «Інваріантність вимірювання» стосується статистичного критерію, що типово вважається необхідним для обґрунтування застосування певної метрики у більше ніж одному культурному контексті. Цей критерій визначає, чи дослідницький інструмент оцінює один і той самий конструкт у різних умовах чи вибірках. Хоча його прихильники наполягають на необхідності інваріантності вимірювання для проведення валідних порівнянь із багатьма групами, його критики піддають його передумови сумніву на підставі того, що вимоги до інваріантності вимірювання є надто суворими, мають виражено заборонний характер чи є хибними. Ландшафт дебатів щодо інваріантності вимірювання є за своєю суттю непростим. Ця публікація систематизує сучасні погляди на інваріантність вимірювання з метою надання історіографічної перспективи на утворення різних позицій у дискусії щодо інваріантності вимірювання у соціальних науках. Розкриваючи складні шари, що оточують проблематику інваріантності вимірювання, а також їх критику та дотичні рекомендації, цей огляд висвітлює ключові розрізнення у панівних підходах та досліджує їхній вплив на постання поточних вимог до інваріантності вимірювання у емпіричних дослідженнях. Ця дискусія має велике значення для викладання дослідницьких методів; набутками обговорення цього питання можна скористатися для покращення та збагачення викладання дослідницького дизайну та методів.
Ключові слова: *вимірювання, інваріантність вимірювання, валідність, кількісні методи, культура, крос-культурні дослідження.*

Relevance and research problem. Social scientists and researchers from adjacent disciplines are often interested in comparing social and cultural groups. The issue of measurement in cross-cultural research continues to give rise to lively and complex discussions, contributing to the examination of the foundational assumptions and practices of psychometrics [20]. The problem at the core of this discussion is whether we can meaningfully and legitimately compare sociocultural or psychological constructs such as personality traits or values across different societies. Albeit with an expansion of culture-relevant research cultural influences in psychological processes are broadly recognized [14] and a strictly universalist approach is relatively rare in the social sciences at present [16; 23], there is still a range of approaches spanning from universalist to relativist views of cultural impact on measurement outcomes. While the universalists consider culture to have universal properties and thus find cross-

cultural comparisons quite feasible – provided the standardized research instruments are used (for example, survey items developed by Geert Hofstede or Shalom Schwartz), – the relativist perspective (which typically emphasizes the importance of capturing the culture-specific features) deems such constructs bound to particular cultures, which, in their view, limits generalizability and undermines the claim of validity in cross-cultural comparisons by pronouncing them practically difficult or even impossible. The third, more middle-ground perspective seeks to reconcile the universalists and relativists, by offering options of using mixed methods or developing culturally sensitive metrics. Finally, the fourth perspective focuses on *measurement equivalence*. Its core idea is that measurement across cultures is possible, provided that equivalence is established and the metrics used in each sample are characterized by invariance.

The criterion of *measurement invariance* that is at the center of the present publication stems from measurement equivalence [16]. The issue of measurement invariance itself has generated a large literature [1; 2; 5; 9; 11; 13; 18; 22; 24; 25] that spans beyond the social sciences [4; 12]. Despite the importance of measurement invariance for validity of findings in cross-cultural research, there is no unanimity in the perceptions of standards for measurement invariance testing. In fact, there has been a long-standing debate regarding measurement invariance which centers around the effectiveness of meaningfully comparing social, cultural and psychological constructs across different sociocultural groups [10; 15]. Indeed, the theme of validity in cross-cultural comparisons is contingent on evidence of comparability of metrics and data. However, despite the growing recognition of the need for cross-cultural validation to compare values, norms and similar traits cross-nationally, there is a relentless tension between the proponents of measurement invariance and its opponents frustrated with the incongruence between the stringent standards and ambiguous recommendations [10; 15]. Practical matters such as losing valuable culture-specific scale items due to imperfect item-total correlations, or finding much smaller than expected psychological differences across cultures are some of the concerns [10; 15]. Frustrating for publishing authors and practitioners and rather confusing for students and young scholars attempting to find the reliable "middle ground" and adhere to the "best practices", the discussion behind the debate appears too entangled to be easily resolved – partly due to the diversity of the perspectives, and partly due to the different purposes different practices of validation have been designed to serve. Meanwhile, the application of measurement invariance techniques remains fragmented and controversial [15]. This review seeks to unpack the layers of complexity surrounding the issue of measurement invariance, tracing their historical development to reveal how their interaction has become the source of the ongoing debate. By peeling off these layers one by one, this review highlights critical distinctions in the prevailing perspectives and examines their implications for the current requirements for measurement invariance in empirical research.

Review of current research and publications. "Measurement invariance" refers to statistical criterion typically considered necessary for using a metric in more than one cultural context [10]. It indicates whether or not an instrument assesses the same construct across different conditions or samples [15]. It is probably fair to suggest that measurement invariance topic, despite its remaining a pivotal concern in cross-cultural research, is rarely met with much enthusiasm among researchers. Many practitioners find the procedures required for demonstration of invariance too convoluted or lacking in substance; some recount 'war stories' about the obstructionist role such requirements played in the dissemination of so painstakingly obtained data from multi-sited ethnographic projects. Although its advocates insist on the necessity of measurement invariance for conducting valid multigroup comparisons, its critics cast doubt on the underlying premises on the grounds of measurement invariance requirements being too stringent, prohibitionist, or misguided [10; 15]. The landscape of measurement invariance debate is inherently complex, yet it is easy, from a cursory engagement, to glean a message that any lack of equivalence is "psychometric ignorance" and any attempts to use metrics that are not invariant are "fatally flawed". One of the downsides of this message is that it tends to discourage researchers from entering or continuing cross-cultural work [10]. Literature in favor of less prohibitionist tone of measurement invariance also points to the generated nature of data used to design recommended stringent invariance procedures, as opposed to the real-life (and imperfect) cross-cultural data [10, p. 890]. Instead, the emphasis on external validity is encouraged (as internal validity does not guaranty external validity), to reach out to real-world conditions, to avoid imposing the isomorphic similarity between the individual and cross-cultural differences [10, p. 893]. A more customized approach, calibrated to the specific needs of the study is advocated instead of blanket use of measurement invariance [15].

Although the notion of measurement invariance is an important benchmark in cross-cultural research, there is some debate regarding its application, standards of testing, and reporting. To make the terrain of this debate more intellectually navigable, **this publication aims to** highlight and critically examine key perspectives with respect to assessment of measurement invariance, evaluate applicable methodological practices, and review relevant recommendations grounded in current literature. Its educational objective is to (1) enhance teaching research design by deepening understanding of the foundational assumptions underpinning measurement

invariance, (2) trace the theoretical evolution of these assumptions, and (3) elucidate the impact of cultural factors on survey research in cross-cultural setting, thereby supporting the development of cross-cultural research competencies.

Divergent perspectives on measurement invariance: Trajectories, critique, recommendations There exists a spectrum of views regarding the appropriate level of stringency for enforcing measurement invariance. Historically, the motivation to establish measurement invariance emerged in the context of ensuring fair selection of candidates for positions [15, p. 3]; therefore the question of high validity standards due to its high stakes was a valuable one. For the purpose of employee selection, the group-level (including cultural) differences were assumed to be irrelevant and the application of strict invariance procedure is understandable as they were developed in response to social pressures and profound social responsibilities [15, p. 3]. On the other hand, psychologists and other social scientists whose interests involve cross-cultural comparisons, have a different agenda with comparing individuals, as for them collective-level attributes are not mere sources of bias or 'noise'. This could be argued to be one of the sources of current divergent perspectives on measurement invariance standards, grounded in different purposes for which measurement invariance may be established. Cross-cultural comparisons are legitimized by the equivalence of metrics used to assess some attribute of interest. If metrics are inequivalent, the comparison across cultural samples is made more difficult to draw conclusions from. Lack of equivalence, in the most general terms, ensues from bias in the data. Biased data lack overlap in the meaning across groups and are not suitable for cross-cultural comparison. The key to measurement invariance in crops-cultural setting is not achieving the absolute invariance but rather ensuring that the irrelevant forms of bias are minimized [15]. The purpose of invariance testing in cross-cultural setting is to establish equivalence across conditions "*so long as those varied conditions are irrelevant to the attribute being measured*" [15].

In light of these nuances, the varied degrees of acceptance for strict procedures within the invariance framework become more systematically structured and comprehensible. As is attested by the literature, there is a spectrum of perspectives ranging from stringent requirement for invariance to methodological re-evaluation [10; 21 for review]. For example, a traditionalist view of measurement invariance dictates that invariance is essential for valid comparisons, which its proponents envision as the way to avoid distortions in conclusions about group differences. Its historiographic roots are in the evolution of the psychometric theory and philosophy of science where invariance is closely aligned with objectivity and measurement is expected to be independent of the sample, behavioral conditions or cultural context [8]. Key psychometric theorists who are credited for laying the foundation for strict invariance include G. Rasch, S.S. Stevens, E.L. Thorndike, and L.L. Thurstone [7]. There are three traditions that gave rise and guidance to strict invariance approach: test score tradition which is concerned with maximized reliability and reduced error variance (e.g. Classical Test Theory); scaling tradition emphasizing designing invariant scales (e.g. Rasch models, Guttman scaling etc.); and structural tradition concerned with modeling latent variables and structural relations (e.g. confirmatory factor analysis). Strict invariance approach draws heavily on the scaling and structural traditions which is reflected in the procedures recommended for testing invariance (e.g., multiple group confirmatory factor analysis (MGCFA) where one checks factor loadings, intercepts and residual variances across different groups or conditions) (for example: https://dlab.berkeley.edu/news/testing-measurement-invariance-using-lavaan-r). This approach also has its critics, who claim that invariance principle is overstated and unnecessarily prohibitionist, while violations of invariance are often inconsequential methods artifact rather than actual validity threats. When taken too far, strict invariance testing can misrepresent the nature of construct studied in cross-cultural comparison.

An alternative, more reform-oriented perspective views approximation to invariance as more realistic and practically achievable than strict (complete) invariance, and allows more flexibility in order to balance the statistical rigor with practical applicability in large-scale cross-cultural research [18; 24; 25]. It stems from hands-on experience with large cross-cultural data sets and calls for a more customized use of measurement invariance protocols that would accommodate the specific research needs. Among the recommendations is the conceptual shift of emphasis to external [10] and nomological [17] validity. For example, construct validity could be enforced by both measurement invariance and theoretical network alignment [for example, 17]. As measurement invariance procedures are often characterized as overly complex, an additional attraction of appealing to nomological validity lies within multiple ways of its demonstration. While structural equation modeling (SEM) is often a default option, several statistical sources [for example, 3] affirm that using correlation- and regression-based analysis is also valid for establishing nomological validity. It also advocates a conceptual shift paying more attention to reflective vs. formative constructs which is essential for valid conclusions in SEM. A more reflexive use of measurement invariance suggests viewing measurement as contextual, as over-reliance on statistical tests may obscure some meaningful and interesting cultural differences. Therefore, rather than uphold rigid statistical thresholds, construct validity, external validity and interpretative caution should be emphasized.

Conclusions. As a hallmark of modern science, measurement is important for research practices and is critical for an efficient investigation in any social

context [3; 6; 19; 26]. In the social sciences measurement often involves human characteristics. In social settings the process of measurement can be understood as connecting unobservable theoretical constructs to observable empirical indicants that can be directly measured. This linkage may blur the distinction between metrological and statistical (data modeling) paradigms in social science, potentially inviting further criticism [20]. Social research is typically dealing with theoretical metrics, i.e. latent parameters that represent the generic and abstract constructs underlying an informant's response to a specific survey item and are by themselves more informative than this response [3; 26]. Cross-cultural research is struggling with methodological pitfalls that need to be addressed by collecting data from different cultural samples [16]. While the standards for measurement invariance have intrinsic value for cross-cultural comparisons, there is still a complex diversity of perspectives and arguments regarding how validity criteria should be met and reported. As validity signals applicability of research findings and their authenticity in the real world, developing valid measurement and constructing instruments that allow adequate coverage of the constructs under study permits claiming data reliability and legitimizes drawing conclusions.

BIBLIOGRAPHY:

1. Boehnke K. Let's compare apples and oranges! A plea to demystify measurement equivalence. American Psychologist. 2022. Vol. 77, № 9. P. 1160–1168. https://doi.org/10.1037/amp0001080

2. Boer D., Hanke K., He J. On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. Journal of Cross-Cultural Psychology. 2018. Vol. 49, № 5. P. 713–734. https://doi.org/10.1177/0022022117749042

3. Carmines E.G., Zeller R.A. *Reliability and validity assessment*. Newbury Park, CA: Sage, 1979. 72 p.

4. Chan K.T., Nguyen T.H., Tran T.V. *Developing cross-cultural measurement in social work research and evaluation*. Oxford: Oxford University Press, 2018. 264 p.

5. De Los Reyes A., Tyrell F.A., Watts A.L., Asmundson G.J.G. Conceptual, methodological, and measurement factors that disqualify use of measurement invariance techniques to detect informant discrepancies in youth mental health assessments. Frontiers in Psychology. 2022. Vol. 13. № 931296. https://doi.org/10.3389/fpsyg.2022.931296

6. DeVellis R.F. *Scale Development: Theory and applications*. (2nd ed.). Thousand Oaks: SAGE Publications, 2003. 171 p.

7. Engelhard G., Jr. Historical views of the concept of invariance and measurement theory in the behavioral sciences. Paper presented at the Fifth International Objective Measurement Workshop, University of California, Berkeley (1989, March).

8. Engelhardt G., Jr. Invariance and the three traditions of measurement in the human sciences. Q. Zhang (ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium* (PROMS 2023) Atlantis Highlights in Social Sciences, Education and Humanities. 2024. Vol. 23. P. 1–13. https://doi.org/10.2991/978-94-6463-494-5_2

9. Fischer R., Karl J.A., Luczak-Roesch M., Hartle L. Why we need to rethink measurement invariance: The role of measurement invariance for cross-cultural research. Cross-Cultural Research. 2025. Vol. 59, № 2. P. 147–179. https://doi.org/10.1177/10693971241312459

10. Funder D.C., Gardiner G. MIsgivings about measurement invariance. European Journal of Personality. 2024. Vol. 38, № 6. P. 889–895. https://doi.org/10.1177/08902070241228338

11. Gerlach P., Eriksson K. Measuring cultural dimensions: External validity and internal consistency of Hofstede's VSM 2013 Scales. Frontiers in Psychology. 2021. Vol. 12. № 662604. https://doi.org/10.3389/fpsyg.2021.662604

12. Herk H.V., Goldman S.P.K. The advancement of measurement invariance testing in cross-cultural research in the period 1999–2020. Executing rather than scrutinizing? *Measurement in Marketing* (pp. 95–119). H. Baumgartner, & B.Weijters (Eds.), (Review of Marketing Research; Vol. 19). Emerald Publishing. 2022. https://doi.org/10.1108/s1548-643520220000019005

13. Iurino K., Saucier G. Testing measurement invariance of the moral foundations questionnaire across 27 countries. Assessment. 2019. Vol. 27, № 2. P. 365–372. https://doi.org/10.1177/1073191118817916

14. Kashima Y. Culture and psychology in the 21st century: Conceptions of culture and person for psychology revisited. *Journal of Cross-Cultural Psychology*. 2015. Vol. 47. № 1. P. 4–20. https://doi.org/10.1177/0022022115599445

15. Kusano K., Napier J.L., Jost J.T. The mismeasure of culture: Why measurement invariance is rarely appropriate for comparative research in psychology. Personality & Social Psychology Bulletin. 2025. Advance online publication. № 1461672251341402. https://doi.org/10.1177/01461672251341402

16. Lacko D., Čeněk J., Točík J., Avsec A., Đorđević V., Genc A., Haka F., Šakotić-Kurbalija J., Mohorić T., Neziri I., Subotić S. The necessity of testing measurement invariance in cross-cultural research: Potential bias in cross-cultural comparisons with individualism– collectivism self-report scales. *Cross-Cultural Research*. 2022. Vol. 56. № 2-3. P. 228–267. https://doi.org/10.1177/10693971211068971

17. Lakshman C., Vo L.C., Ramaswami A. Measurement invariance and nomological validity of the Attributional Complexity Scale: Evidence from Estonia, France, India, United States, and Vietnam. *International Journal of Cross Cultural Management*. 2020 Vol. 20. № 1. P. 89–111. https://doi.org/10.1177/1470595820913600

18. Robitzsch A., Lüdtke O. Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons. Structural Equation Modeling. 2023. Vol. 30, № 6. P. 859–870. https://doi.org/10.1080/10705511.2023.2191292

19. Tal E. Measurement in Science. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.). 2020. https://plato.stanford.edu/archives/fall2020/entries/measurement-science

20. Uher J. Statistics is not measurement: The inbuilt semantics of psychometric scales and language-based

models obscures crucial epistemic differences. *Frontiers in Psychology*. 2025. Vol. 16. 1534270. https://doi.org/10.3389/fpsyg.2025.1534270

21. Van De Schoot R., Schmidt P., De Beuckelaer A., Lek K., Zondervan-Zwijnenburg M. Editorial: Measurement invariance. Frontiers in Psychology. 2015. Vol. 6, Article 1064. https://doi.org/10.3389/fpsyg.2015.01064

22. Vesper D., König C.J., Siegel R., Friese M. Is use of the general system justification scale across countries justified? Testing its measurement equivalence. The British Journal of Social Psychology. 2022. Vol. 61, № 3. P. 1032–1049. https://doi.org/10.1111/bjso.12520.

23. Wang Q. Why should we all be cultural psychologists? Lessons from the study of social cognition. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science.* 2016. Vol. 11. № 5. P. 583–596. https://doi.org/10.1177/1745691616645552

24. Welzel C., Inglehart R.F. Misconceptions of measurement equivalence: Time for a paradigm shift. Comparative Political Studies. 2016. Vol. 49, № 8. P. 1068–1094. https://doi.org/10.1177/0010414016628275

25. Welzel C., Brunkert L., Kruse S., Inglehart R. F. Non-invariance? An overstated problem with misconceived causes. Sociological Methods & Research. 2023. Vol. 52, № 3. P. 1368–1400. https://doi.org/10.1177/00491241219955 21

26. Zeller R.A., Carmines E.G. *Measurement in the social sciences: The link between theory and data*. Cambridge: Cambridge University Press, 1980. 198 p.