Critical Revision of the Mental Health Questionnaire /SUSESO CEAL-SM

Claudio Garuti
claudio@medicalsapiens.cl

Mario Sandoval
mario@medicalsapiens.cl

Highlights

- Critical view of the international manual CEAL SM/ SUSESO
- Metric methodology (cardinal vs ordinal)

ABSTRACT

The Work Environment Assessment Questionnaire – Mental Health / SUSESO, CEAL-SM/SUSESO, is established as an instrument for identifying and measuring psychosocial risk factors at work and, according to its website, claims to be a measurement instrument. This questionnaire consists of 88 questions that are grouped into two sections: a general one, with 34 questions; and a psychosocial risk one, with 54 questions. The risk exposure score or "risk status" is only calculated for one specific section and, by definition, is the raw sum of the scores obtained in each of the sections of this specific section. Those scores are based in ordinal scales; thus the results are lack of rationality. Because of this, the user obtains unsure outputs that may not reflect the measured environment.

Keywords: Multicriteria Analysis, AHP/ANP, Scale of Measurement, Threshold, ordinal-cardinal scales.

Introduction

The Work Environment Assessment Questionnaire – Mental Health / SUSESO, CEAL-SM/SUSESO, is established as an instrument for identifying and measuring psychosocial risk factors at work and, according to its website, claims to be a measurement instrument. This questionnaire consists of 88 questions that are grouped into two sections: a general one, with 34 questions; and a psychosocial risk one, with 54 questions. The risk exposure score or "risk status" is only calculated for one specific section and, by definition, is the raw sum of the scores obtained in each of the sections of this specific section. Those scores are based in ordinal scales; thus the results are lack of rationality. Because of this, the user obtains unsure outputs that may not reflect the measured environment.

2. Literature Review (State of Art)

CEAL-SM/SUSESO is based in The Copenhagen Psychological Questionnaire COPSOQ. First version of this questionnaire was published in 2005[1]. After, the COPSOQ has had several modifications, the last was COPSOQ III in 2021, review in 2023[2]. In Chile COPSOQ III was evaluated by SUSESO authorities, tested and validated[3]. After that, SUSESO decided applied to all work populations.
To measure Mental Health SUSESO choice General Health Questionnaire -12 (GHQ-12), created in 1972 by Golberg. Its was translate to many languages and has several criticizes about its application

and synthesized. In some cases, researchers had use factorization u others methods to fix or adapt some results, by has 12 questions with 4 kinds of answer where only exist 2 options 0 or 1, and all question have the same value[4].

3. Objectives

Carry out a critical review of the method used to create the COPSOQ and the new version SUSESO CEAL-SM mental health questionnaire.

4. About The Methodology

One of the main virtues of this work is the disaggregation of the dimensions into questions that explain them, facilitating the understanding and comprehension of what is being sought, that is, structuring the problem into its components.

However, the questionnaire presents two fundamental errors when trying to generate a metric:

1. The lack of priorities between the questions in the questionnaire (every question is equally important) and

2. A synthesis of results without solid mathematical foundations.

Regarding the first, it seems, based on how the calculations are carried out, that the dimension treated (and therefore the questions defined by it), always have the same importance and are indifferent to their origin, which seems very unintuitive, lacking a rational explanation in this regard.

Regarding the second, a transversal application of a structural type can be observed on all the criteria of the model of the same type of scale. For example, to answer each question in the specific section, the same Likert scale format is used: Always, Often, Sometimes, Rarely and Never/Almost never, where each of these answers is equivalent to Always (0), Often (1), Sometimes (2), Rarely (3) and Never/Almost never (4). Only for the Vulnerability Dimension the scale is from 1 to 5. The rest start from 0 (to avoid "adding" risk where it does not correspond).

The work attempts to generate a metric for measuring the state of risk. By definition the concept of metric is: the measure that serves to quantify the proximity or distance between two elements of a geometry in a space measurement. For example, given two elements A and B where element A is 2 units above the ground and element B is 6 units above the ground, two calculations can be performed:

1. Subtraction: The distance between elements A and B is 4 units (6 – 2 = 4)

2. Proportion or ratio: Element B is 3 times farther from the ground than element A (6/2 =3).

These mathematical relationships, which are fundamental to creating a metric, cannot be carried out when using an ordinal scale such as the Likert scale, since its numbers are only a classification of order and do not reflect, under any point of view, relative distances or ratio between the elements. The scale used in this work (almost the only one) is: *Always (0), Often (1), Sometimes (2), Rarely (3) and Never/Almost never (4)*, which is a linear scale. But, is the real world linear? Also, how to calculate (and justify) that level *Rarely* is 3 times worse than level *Often*? Who said so? This is an ordinal type of scale. Therefore, any arithmetic operation (addition, multiplication, subtraction or division) in this Likert scale is not defined (not permitted), and its zero to represent the absence of the property is also not defined (have not even sense in an ordinal scale).

5. About The Results

Unfortunately, it is very common to see that in professional work people continue to believe that a classification, just because it is given in numbers (instead of letters or any other symbol), gives rise to performing arithmetic operations with them. It is even common to see how some people normalize these numbers (for example, transforming 1, 2, 3, 4 in 25%, 50%, 75% and 100%), thinking that in this way it is given the necessary mathematical attributes to operate with them.

Another precept equivalent to the previous one and also observed in the questionnaire, is the mere fact that a question elicits a greater number of negative answers (simple counting), implies that it is the most important to solve the problem. It should be said clearly and without mincing words that counting is not equivalent to measuring (quantity is not synonym of quality neither frequency synonym of importance).

Finally, regarding the delicate point of thresholds or tipping points (at what number of points the risk score becomes unacceptable). In this work, the threshold is determined in a purely empirical way, that is, the scores obtained for a certain population are compared in the field and according to their behavior threshold is established in a completely separate way from the constructed metric. This is very dangerous to do because it may depend on variables that are not integrated in the metric (saying nothing of the metric was badly constructed). Strictly speaking, in the best case, this procedure can only be used as a validator of results, not as a predictor of them.

They also use another discussible way to calculate thresholds, they trisect (divide in 3 equal parts) the final rating with the first part being the riskiest (red flag), the middle part the moderate risk (yellow flag) and the lower part the one with lower risk or without risk. Again, this procedure has no mathematical explanation and it seem just an *easy way to come out*.

6. Conclusions

The rationality of the questionnaire suffers from several basic flaws, such as the importance of the questions, the scale used, the operations allowed and the threshold calculation.

The above leaves many open questions as “what happens if”, regarding the results obtained. We mean, what happen if some important criteria in the field are not present in the methodology? Why the criteria are equal important and the risk score just depend of the number (the quantity) of responses instead of their importance? Why the scale is unique and ordinal class? And finally, why the threshold (a very relevant issue) is obtained without any mathematic rationality?

The above questions need to be answered to give top credibility to the Manual. Thus, it is necessary to correct these flaws, to give the questionnaire adequate rigor and then the necessary flexibility to adjust it based on the environment where it is applied, without losing rationality in the process.

7. Limitations

This paper is a critical view of the international manual CEAL SM/ SUSESO. Thus, the Limitations issue has no place.

8. Key References

1. Kristensen T, Hannerz H, Høgh A, Borg V. The Copenhagen Psychosocial Questionnaire—a tool for the assessment and improvement of the psychosocial work environment. Scand J Work Environ Health. 2005;31(6):438–449.
2. COPSOQ international network. COPSOQ III. Guidelines and questionnaire. [online].; 2021 [Reviewed in February 2023. Available in: "https://www.copsoqnetwork.org/assets/Uploads/COPSOQ-network-guidelines-an-questionnaire-COPSOQ-III180821.pdf".
3. Nuevo cuestionario de evaluación de riesgo psicosocial CEAL-SM / SUSESO Proceso de validación de un nuevo instrumento para el diagnóstico y la medición de los factores de riesgo psicosocial en Chile. Juan M Pérez-Franco, Macarena Candia Tapia, Marco Soto. Equipo de Riesgo Psicosocial Laboral. Unidad de Prevención y Vigilancia. Intendencia de Seguridad y Salud en el Trabajo. Documento trabajo N°24. https://www.suseso.cl/607/articles-703537_archivo_01.pdf.
4. The 12-Item General Health Questionnaire (GHQ-12): translation and adaptation study of the Romanian version. Andrea C Brabette. Revista Iberoamericana de Diagnóstico y Evaluación Psicológica. ISSN: 1135-3848. N°37-Vol 1-2014.
5. WEBPAGE SUCESO. https://www.suseso.cl/606/w3-propertyvalue-614691.html#presentacion.
6. Garuti, C. Mu E. (2024). A rate of change and center of gravity approach to calculating composite indicator threshold: Moving from an empirical to a theoretical perspective. Mathematics 2024, Volume 12, Issue 13, 2019.
7. Garuti, C. (2016). Consistency & Compatibility. Two sides of the same coin (but with different implications). ISAHP 2016 International Symposium of the Analytic Hierarchy Process.
8. Garuti, C. (2014). Measuring in Weighted Environments. International Journal of AHP (IJAHP) 2014.
9. Garuti, C. (2012). Measuring in Weighted Environments: Moving from Metric to Order Topology. Santiago, Chile: Universidad Federico Santa Maria (Book, 72pp).
10. Garuti, C. (2014). Compatibility of AHP/ANP vectors with known results. Presentation of a suggested new index of compatibility in weighted environments. International Symposium of the Analytic Hierarchy Process.
11. Garuti, C. (2007). Measuring Compatibility in Weighted Environments: When close really means close? International Symposium on AHP, 9, Viña del Mar, Chile. 2007.
12. Garuti C. 2021, "How to Obtain a Global Reference Threshold in AHP/ANP". IJAHP International Journal of AHP Vol. 13. N°1. Abstract 533 | PDF Downloads 166 | DOI https://doi.org/10.13033/ijahp.v13i1.802
13. Garuti C. 2022, "A Set Theory Justification of Garuti´s Compatibility Index: generalization of Jaccard index working within a weighted environment". Advances in Mathematics Research. Vol.30 e-ISBN: 978-1-68507-536-1. DOI: https://doi.org/10.52305/BDFQ3979. Editor: Albert R. Baswell. Published by Nova Science Publishers NY.
14. Saaty, T.L., 1994, Fundamentals of Decision Making and Priority Theory. McGraw Hill Publishers; New York. RWS Publications. ISBN-10: 0-9620317-6-3. ISBN-13: 978-0-9620317-6-2.
15. Sandoval M. and Garuti C., May 2006. The AHP: A multicriteria decision making methodology for shiftwork prioritizing. Journal of Systems Science and Systems Engineering 15(2):189-200. DOI:10.1007/s11518-006-5007-5

16. EVA SCHÖNFELDER & PETER KNAUTH (1993) A procedure to assess shift systems based on ergonomic criteria, Ergonomics, 36:1-3, 65-76, DOI: 10.1080/0014013930896785: http://dx.doi.org/10.1080/00140139308967856