Outbreak, Surveillance, Investigation & Response (OSIR) Journal
Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, http://www.osirjournal.net

The Grammar of Science: Do You Get What You Measured?

Jaranit Kaewkungwal

Mahidol University, Thailand

Corresponding author email: jaranitk@biophics.org

Received: 25 May 2023; Revised: 20 Jun 2023; Accepted: 29 Jun 2023
https://doi.org/10.59096/osir.v16i2.263899

Old sayings—"You get what you measured.", "What gets measured gets done/managed.", "You can't improve what you don't measure.", etc. So…when conducting a research study, are you sure that the research instrument measure what it is intended to?

Research instrument is a tool used to collect, measure, and analyze data related to the research question of interest. Designing or choosing an instrument is the initial process after conceptualization of research questions and operationalizing concepts or constructs of the instrument.[1] When selecting or developing a research instrument, researchers must be assured that the instrument possesses quality characteristics including its reliability and validity, basing on conceptual framework of the study. Research tools can be in any formats: questionnaire, interview form, direct-indirect observations laboratory test, etc. We will now focus on validity of questionnaire or test battery designed to be administered in order to obtain a comprehensive assessment of a particular factor or phenomenon.

Validity of Research Instrument

There are three types of validity: construct validity, content validity and criterion validity. Construct validity is appraised when researchers desire to draw an inference from scores (obtained from the questionnaire or test) that can be grouped under the label of a particular construct while content validity assesses whether the items in the questionnaire or test adequately represent each specific construct of interest. Criterion validity is usually done when researchers want to draw an inference whether or not scores obtained from the new developed measurement/test are associated with, or could predict, other observable variable that has practical importance.[2–5]

As an example, when designing a questionnaire to measure quality of life (QOL), researchers may perform in different steps. The process usually starts with identifying and defining the constructs of QOL based on hypothetical or theoretical concept(s) of QOL, say—physical function, psychological function, social function, and spiritual function.[6] Next step is writing up the elements (questions or times) for the questionnaire by taking into consideration of "content validity" in terms of item relevancy, representativeness, necessity, and completion of the concept(s) supposed to be measuring. In order to ensure the "construct validity", the questionnaire should be assessed whether it actually covers and measures the four desired constructs of QOL. Moreover, the measurement scores obtained from the questionnaire could be tested for "criterion validity" ensuring that it is theoretically and logically accurate with the desired constructs by associating the scores with a certain criterion concurrently measured (concurrent validity) or making prediction with a certain outcome in the future (predictive validity); e.g., QOL scores may be correlated with health perception at present time and health improvement status thereafter.

Construct validity can be viewed as an overarching term to assess the validity of a research instrument as it incorporates other forms of validity (i.e., content validity, and criterion validity).[5] That means confirming construct validity covers the process of testing content validity, criterion validity and others.

Types of Construct Validity: Convergent Validity & Discriminant Validity

Construct validity is one of the most important concepts in measurement theory, especially when researchers want to measure a variable that is not itself directly observable. Many times, the variables to be measured in a study is not concrete or obviously observable. Not only research in social sciences, psychology and education, but also those in medicine

and public health measure and analyze abstract concepts; for example, pain, anxiety, quality of life, and satisfaction to a healthcare program.[7–10] It is thus important to evaluate whether the research instrument is, in fact, measured the "hidden" or "latent" variables it is supposed to measure or not.

Researchers typically establish construct validity by correlating a measure of a construct with the other measures that should, theoretically, be associated with it (so-called convergent validity) or vary independently of it (so-called discriminant validity or divergent validity).[11] Convergent validity is observed when the constructs expected to be related are shown related. On the other hand, discriminant validity is observed when the constructs that should not be related are, in fact, unrelated.[7,9,12] In appraising construct validity of QOL, the convergent validity would assess that QOL is, in fact, related to a few overlapped concepts, e.g., life satisfaction or self-appraisal. Discriminant validity would ensure that QOL does not related to or has very poor relationship with the non-overlapping factors, e.g., ethical or political ideation.[7,13] That means, the questionnaire or test should possess two qualities; the test should measure "what it is to measure" and does not measure "what it should not be measured". The research instrument that shows such quality is regarded as having excellent construct validity.[7,9]

It should be noted that convergent validity is a type of construct validity and it is not the same as concurrent validity which is a type of criterion validity. Both convergent and concurrent validity are assessed by examining the correlation between the scores from two measurements. Concurrent validity compares a new measurement with a well-established measurement in attempt to present the criterion validity. On the other hand, convergent validity tests the relationship between two new measurements or two related constructs.[3]

Researchers can determine construct validity of a research instrument by different methods. Simple analytical methods include: comparison of the scores obtained from the instrument among known groups (known-groups validity), or correlation with scores from other instrument/test that measures similar qualities or constructs. Other advanced methods include: factor analysis, multitraits-multimethods (MTMM), structural equation modeling (SEM), etc.

Construct Validity & Known-groups Validity

A simple method to provide an evidence of construct validity is to use "known-groups validity" to confirm the constructs. The method is to compare scores obtained from the questionnaire/ test between independent groups (or extreme groups) with vs. without the characteristics related to the constructs being measured. The expectation of the comparison is to be able to detect hypothesized differences between these known independent groups. If scores obtained from the measurement can effectively differentiate between the groups, then the construct validity may be established.[14]

Statistical techniques used to compare different known groups could be Independent samples t-tests, ANOVAs, and non-parametric tests. Researchers can simply assess construct validity by using t-test to compare QOL physical function scores between those who apparently show good physical function vs. those who do not. Here are some examples in literature showing that the researchers might compare QOL scores among different subgroups in several known conditions. A study assessing validity for a QOL instrument employed known-group validity to discriminate among elderly groups known to be different in varying levels of aged-care needs and self-reported health.[15] The other study assessed construct validity of QOL score among elderly groups with different education levels and the presence or absence of the chronic conditions (dementia, depression, and dizziness).[16] Another study compared generic measures of QOL among different subgroups of adolescents from different contexts, subgroups of outpatient treatment for mental disorders and subgroups of students with low and high human development.[17]

Construct Validity & Factor Analysis

Construct validity can be addressed by a factor analysis. Factor analysis is typically used to analyze interrelationships within a set of variables or to confirm the construction of a few hypothetical variables, so-called factors (domains, dimensions, components).[2] Historically, it was noted by measurement experts that "construct validity" has been known as 'factorial validity'.[10,18,19]

Factor analysis answers the question asked by construct validity. In general, the method is to postulate common latent (hidden) factor(s) from the intercorrelations among the observed variables (i.e., items in the questionnaire).[2] There are two types of factor analysis, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).[2,20] Based on the interrelationships among a set of observed variables/items, EFA answers: what are the hidden factors (constructs) and whether they are the factors (constructs) that the questionnaire/test is supposed to be measured. As EFA is the technique to identify the

factor structure or model hidden in a set of observed variables; it is thus considered as a "theory-generating" procedure. On the other hand, CFA answers whether different sets of variables/items are correlated together according to the different hypothetical factors (constructs) that the researchers used as basis when they developed that questionnaire/test. CFA is thus considered as a "theoretical-testing" approach.[2,20–22]

The analytic question of EFA is: "What are the underlying or latent constructs that could have produced from the observed pattern of covariances (or correlations) among the variables/items?" EFA results in factor loadings which are derived from correlations between a set of variables/items and a hypothetical construct. Typically, a minimum number for factor loading ranging between 0.30–0.55 is considered to be a strong factor loading coefficient. Based on EFA results, the constructs (latent variables) are established by deducing names for the factors based on the content (i.e., wording) of the items (observed variables) that load heavily upon them.[22]

The analytic question for CFA is: "Are the covariances (or correlations) among a set of variables/items consistent with a hypothesized factor structure?" That means, CFA confirms the relationships between observed variables/items with factors (constructs) that is generally based on a strong theoretical or observational foundation that researchers used to specify an exact factor structure in advance. In general, CFA restricts which variables/items will load on which factors, as well as which factors will be correlated.[22]

As an example, based on literature review on theories related to QOL, researchers may decide to develop a QOL measurement composing of three main constructs, i.e., physical health, mental health, and social health.[6,22–24] The analysis of EFA examines interrelationships among all items with the three "hidden" constructs while CFA confirms interrelationships among a set of items with the specific pre-planned construct. The three constructs may be independent or may be related to one another depending on the researchers' conceptualization of the QOL constructs (Figure 1).

Several techniques can be selected to perform factor analysis including, for examples, principal axis factoring, principal components analysis, image factor extraction, alpha factoring.[22] Here are some examples in literature that employ factor analysis to determine construct validity. A study determined the construct validity using EFA to show the existence of four dimensions: physical, psychological, social and spiritual domains.[25] In another study, after reviewing several theories of QOL, the researchers employed EFA to identify a parsimonious and psychometrically sound solution factors of QOL and subsequently confirmed the factors using CFA.[21]

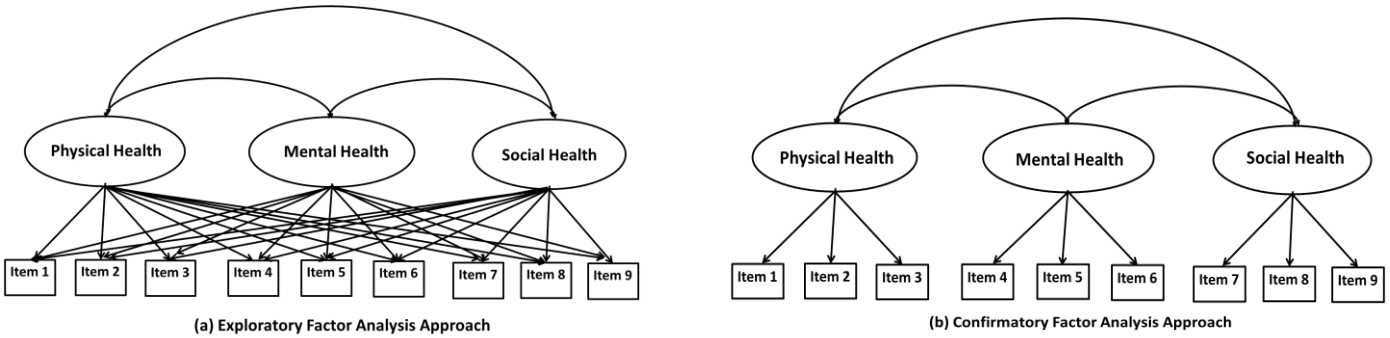

Figure 1. Factor Analysis Models

Construct Validity & Multitraits-multimethods (MTMM)

MTMM was introduced by Campbell and Fiske in 1959.[9] MTMM assesses the common "set of associations" among several traits, each being measured via several methods. Traits could be hypothetical constructs while methods can be different procedures of data collection method, different groups of data provider, or different types of items in the measurement/test.[24,26–28] MTMM evaluates trait-related and method variances through the correlations among scores obtained from multiple constructs and multiple methods.[4] As an example, shown in Figure 2, researchers want to assess correlations among scores of three traits (constructs) of QOL (physical, mental and social) as scaled by two methods (self-rating and caregiver rating).

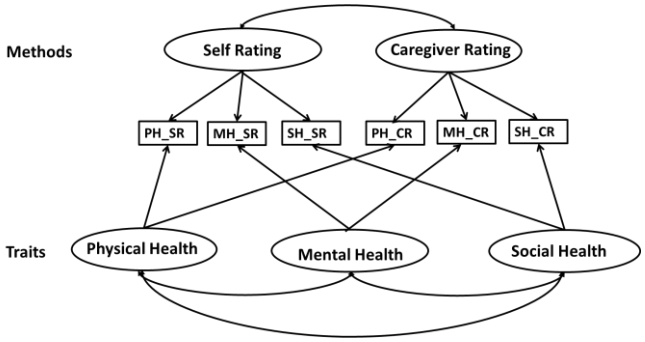

Figure 2. Multitraits-multimethods approach

MTMM involves a correlation matrix customized to enhance the evaluation of construct validity in terms of the discriminant and convergent validity. As shown in Figure 3, MTMM will assess: (1) the relationship between the same constructs (traits) and the same methods (i.e., reliability of the measurement, shown in the diagonal), (2) the relationship between the same constructs (traits) using different measurement methods (i.e., convergent validity); and (3) the relationship between different constructs using different methods of measurement and between different constructs with same method (i.e., discriminant validity). In determining whether the questionnaire/test has a good construct validity or not, the correlations for convergent validity should be high while the correlations for the discriminant validity should be low.[26] As a rule of thumb, an r value of >0.50 is generally considered sufficient to suggest convergent validity. It should also be noted that correlations among related constructs should be higher than those of unrelated constructs.[12]

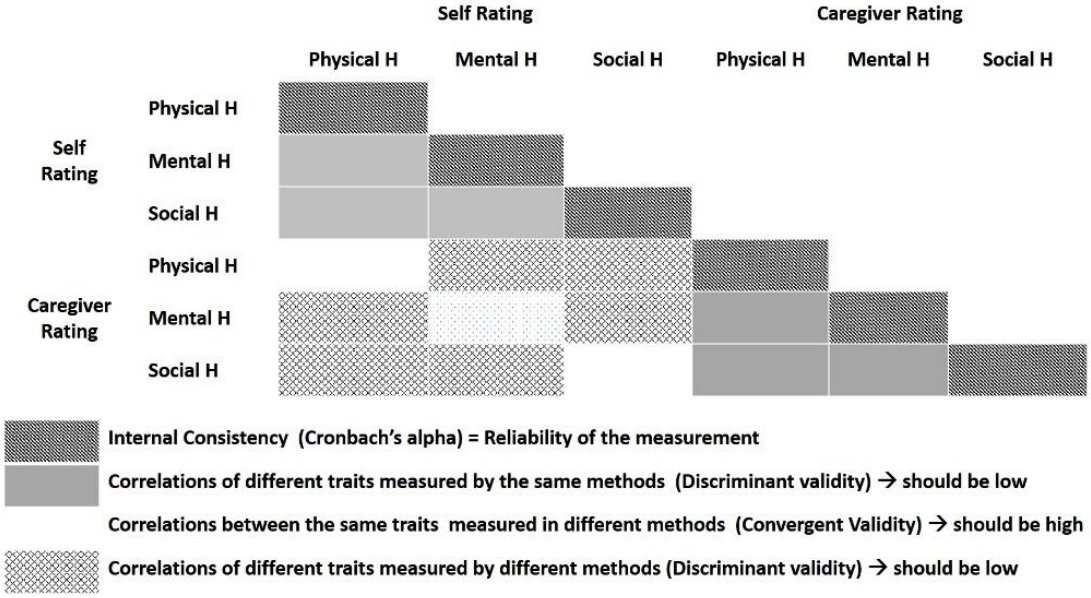

Figure 3. Content validity ratio (CVR) based on number of experts evaluated the item essential

Some examples of MTMM approach used to assess construct validity in literature. A study examined the convergent and discriminant validity of the four domains of the WHO-QOL-BREF by using MTMM approach; each participant filled out the WHO-QOL-BREF questionnaires of four traits (physical, psychological, social and environmental functions) using four different scaling methods (Likert-type scale, visual analogue scale, pie scale, and partner rating).[27] The other study employed MMTM to assess construct validity of children's QOL questionnaire with four subscales/constructs (physical, emotional, social, school functioning) with two methods (child-self report and parent-proxy report.[23] MTMM can be used to assess construct validity of the new research instrument against the constructs of the standard instrument. As an example, a study employed MTMM to assess the convergent and discriminant validity of QOL which composed of three traits/constructs or subdimensions (i.e., physical, social, and psychological well-being) and two methods (i.e., two QOL instruments, standard WHO-QOL and new Roh QOL developed for middle age adult).[24] Another example, in a study on developing a translated and culturally adapted version of QOL questionnaire, the construct validity was first assessed by using EFA and then appraised for the convergent and discriminant validity using MTMM.

The study assessed three domains for QOL (physical functioning, emotional well-being, social functioning) with two methods (i.e., two questionnaires, new translated QOL and standard WHO-QOL).[28]

Conclusion

Researchers should be assured that they really measure what they intend to measure. That is, the questionnaire or other measuring tools developed, adapted or adopted to be used in their study should possess good construct validity. If the instrument lacks construct validity, the study results will be difficult to interpret.[11] To demonstrate construct validity may not be done in one single analysis because it is also related to content validity and criterion validity.[3]

Several factors can be threats to construct validity. The most important aspect of construct validity is about possessing clear and precise conceptual and operational definitions of the constructs intended to measure.[3] Definition of a construct can invalidate the measurement if the definition is incomplete, inexact, unclear, mislabeling, too broad, or too narrow. Construct confounding, which would lead to unreliable scores, may occur when other constructs overrule or mask of the effects of the measured construct.[7] For example, when assessing QOL constructs which could be affected by several factors (e.g., disease severity, social status), and if those factors were not taken into account, one might face construct confounding problem.[3,7]

You may have a great idea and a very good research question, but it is even more critical to ask yourself: "What is the quality of my research instrument?" "Is it reliable and valid?" "Does it measure what I want to measure to answer my great research question?"

Suggested Citation

Kaewkungwal J. The grammar of science: Do you get what you measured? OSIR. 2023 Jun;16(2):113–8.

References

1. Teachers College, Columbia University. Research instrument examples [Internet]. New York: Teachers College, Columbia University; [cited 2023 May 20]. 8 p. <https://www.tc.columbia.edu/media/administration/institutional-review-board-/guide-amp-resources---documents/Published_Study-Material-Examples.pdf>
2. Stapleton CD. Basic concepts in exploratory factor analysis (EFA) as a tool to evaluate score validity: a right-brained approach [Internet]. Arlington (VA): Clearinghouse on Assessment and Evaluation;1997 Jan [cited 2023 May 20]. <http://ericae.net/ft/tamu/Efa.htm>
3. Laerd Dissertation. Construct validity [Internet]. [place unknown]: Lund Research Ltd; [cited 2023 May 20]. <https://dissertation.laerd.com/construct-validity-p4.php>
4. Straussi ME, Smithii GT. Construct Validity: Advances in Theory and Methodology. Annu Rev Clin Psychol. 2009 Apr 27;5:1–25. doi:10.1146/annurev.clinpsy.032408.153639.
5. Messick S. Test validity and the ethics of assessment. Am Psychol. 1980 Nov;35(11): 1012–27. doi:10.1037/0003-066X.35.11.1012.
6. Allen J, Inder KJ, Lewin TJ, et.al. Construct validity of the Assessment of Quality of Life - 6D (AQoL-6D) in community samples. Health Qual Life Outcomes. 2013 Apr 17;11:61. doi:10.1186/1477-7525-11-61.
7. Shuttleworth M. Convergent and discriminant validity [Internet]. [place unknown]: Explorable.com; 2009 Aug 21 [cited 2023 May 20]. <https://explorable.com/convergent-validity>
8. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955 Jul;52(4):281–302. doi:10.1037/h0040957.
9. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56(2): 81–105. doi:10.1037/h0046016.
10. Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill: 1978. 701 p.
11. Westen D, Rosenthal R. Quantifying construct validity: two simple measures. J Pers Soc Psychol. 2003 Mar;84(3):608–18. doi:10.1037//0022-3514.84.3.608.
12. Nikolopoulou K. What is convergent validity? [Internet]. [place unknown]: Scribbr; 2022 Aug 31 [updated 2022 Nov 30, cited 2023 May 20]. <https://www.scribbr.com/methodology/convergent-validity/>
13. Sturm C, Gutenbrunner CM, Egen C, et.al. Which factors have an association to the Quality of Life (QoL) of people with acquired Spinal Cord Injury (SCI)? A cross-sectional explorative observational study. Spinal Cord. 2021;59:925–32. doi:10.1038/s41393-021-00663-z.

14. Heidel E. Known-groups validity: Compare independent groups on survey outcomes [Internet]. [place unknown]: Scalestatistics.com; [cited 2023 May 20]. <https://www.scalestatistics.com/known-groups-validity.html>

15. Khadka J, Ratcliffe J, Hutchinson C, Cleland J, Mulhern B, Lancsar E, et al. Assessing the construct validity of the Quality-of-Life-Aged Care Consumers (QOL-ACC): an aged care-specific quality-of-life measure. Qual Life Res. 2022 Sep;31(9):2849–65. doi:10.1007/s11136-022-03142-x.

16. Hofman CS, Lutomski JE, Boter H, Buurman BM, de Craen AJ, Donders R, et al. Examining the construct and known-group validity of a composite endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS); A large-scale data sharing initiative. PLoS ONE. 2017 Mar 15;12(3): e0173081. doi:10.1371/journal.pone.0173081.

17. Rodrigues S, Pedroso J, Pontes F, Kappler C. Measuring Health-Related Quality of Life in Adolescents by Subgroups of Students and Outpatient Mental Health Clients. Psychology. 2015 Jun;6(7):833–45. doi:10.4236/psych.2015.67082.

18. Anastasi, A. Evolving concepts of test validation. Annu Rev Psychol. 1986;37:1–15. doi:10.1146/annurev.ps.37.020186.000245.

19. Guilford JP. New standards for test evaluation. Educ Psychol Meas. 1946;6:427–439.

20. Matsunaga M. How to factor-analyze your data right: do's, don'ts, and how-to's. International Journal of Psychological Research. 201;3(1):97–110.

21. Reeves AJ, Baker RT, Casanova MP, Cheatham SW, Pickering MA. Examining the factorial validity of the Quality of Life Scale. Health Qual Life Outcomes. 2020 Feb 18;18(1):32. doi:10.1186/s12955-020-01292-5.

22. Swisher LL, Beckstead JW, Bebeau MJ. Factor analysis as a tool for survey analysis using a professional role orientation inventory as an example. Phys Ther. 2004 Sep;84(9):784–99.

23. Bastiaansen D, Koot HM, Ferdinand RF, Verhulst FC. Quality of life in children with psychiatric disorders: self-, parent, and clinician report. J Am Acad Child Adolesc Psychiatry. 2004 Feb;43(2):221-30. doi:10.1097/00004583-200402000-00019.

24. Chung WC, Kim WK, Cha SJ, Kim JR, Kang HH. Convergent and discriminant validity of quality-of-Life questionnaires. Research Quarterly for Exercise and Sport. 2014;85 Suppl 1:A44.

25. Barrera-Ortiz J, Carrillo-Gonzalez GM, Mabel G, Chaparro-Diaz L, Sanchez-Herrera B, Vargas-Rosero E, et.al. Construct validity and reliability of quality of life instrument family version in Spanish. Enfermeria Global. 2015;14(37):239–49.

26. Kyriazos TA. Applied psychometrics: the application of CFA to multitrait-multimethod matrices (CFA-MTMM). Psychology. 2018;9: 2625–48.

27. Hsiao YY, Wu CH, Yao G. Convergent and discriminant validity of the WHOQOLBREF using a multitrait-multimethod approach. Soc Indic Res. 2014 May 1;116(3):971–88. doi:10.1007/S11205-013-0313-Z.

28. Senanayake S, Gunawardena N, Palihawadana P, Jularatna S, Peiris TSG. Validity and reliability of the Sri Lankan version of the kidney disease quality of life questionnaire (KDQOL-SF™). Health and Qual Life Outcomes. 2017 Jun 5;15(1):119. doi:10.1186/s12955-017-0697-6.