

# **Confirmatory Factor Analysis to evaluate the Construct Validity of a Questionnaire: A Survey form developed to assess the Data Governance Performance of a healthcare entity**

*Sara Alshaya, Aji Gopakumar, Santosh Nair*

## **RELEVANCE:**

Staff Survey is a frequently used instrument designed for different purposes in an organization. The survey method includes the process of developing a form including multiple variables, constructs and other personal details. The content of the survey will be organized with multidimensional constructs or latent variables under which multiple observed variables will be listed. The proposed study uses a set of statistical techniques to measure and analyze the relationships of these observed and latent variables. A structural equation modeling (SEM) is proposed to analyze the structural relationship between measured variables and latent constructs which is a combination of factor analysis and multiple regression analysis. The model is applied to validate a newly created survey form by finding linear causal relationships among the variables with measurement error. The study integrates best features of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), that perform construct validity and provide confirmatory tests to validate the specified constructs proposed for the survey form. The research proposes a model and fitness of the model which will be identified to justify the inclusion or removal of the relevant variables/constructs in the survey form.

## **CONTEXT AND AIMS:**

There are several validated questionnaires for different purposes according to the objectives and goals that are need to be achieved. However, many forms available in the literature are reliable and readily usable, the current study developed a new form to assess the staff perception/opinion regarding the data governance performance in the entity. The organization has a centralized unit named Data and statistics department (DSD) which is in the stage of creating a data governance framework with a strong vision, mission and long-term goals. This is to enhance data related operations of the entity with integrity, confidentiality, consistency and accessibility across its each process and procedures. Data and reports generated by the unit is bound by pre-defined rules, and unified standards with clearly assigned ownership criteria. The statistics and reports shared by the unit help the stakeholders to take confident decisions based on its trustworthy data assets. As the unit has an outcome metrics for which measuring the impact is an important step and it is essential for establishing a data culture in EHS to have successful patient outcomes. In this context, a survey form has been developed to assess the performance at different levels and it is to be validated to identify the extent to which a measurement instrument is capable of measuring what it is supposed to measure. Therefore, the aim of the current study is to perform an exploratory and confirmatory factor analysis for ensuring construct validity of the form. In order to apply CFA, the study assumed a priori model specified from the EFA.

## **METHODS AND FINDINGS:**

DSD has developed a survey form including the rules, processes, procedures and responsibilities to be handled in the unit to ensure the ethics and compliance to the data management guidelines of the organization. The survey instrument includes 30 items and EFA was applied in a data of pilot survey to assess and validate the data governance framework developed by the department (DSD). The validation process helps to find the potential constructs or latent variables towards the overall concept of 'staff perception regarding DSD functions'. As a preliminary step, the potential constructs and its respective items were identified from the EFA which were tested in the stage of CFA. The constructs identified are validated and confirmed with CFA which offered a measurement model based on structural equation modeling. This measurement model quantitatively defined the relationships between observed variables (items) and unobservable concepts (factors / latent variables).

Reliability analysis resulted with a reliability coefficient (Cronbach's alpha value) of 0.98 as the measure of internal consistency which concluded that items are closely related to measure the overall construct 'staff perception on data governance performance of DSD'. In the EFA, there was no multicollinearity identified since the determinant of the correlation matrix is greater than 0.00001. principle component analysis was used as the extraction method. Item communality ranges from 0.67 to 0.97 (ideal) with an average variance 0.87 which is being captured by the factor model. This exploratory factor analytic (EFA) study has found a three-factor (process improvement, people empowerment, technology implementation) structure for this instrument; the CFA was developed with the constructs identified from EFA. CFA predicted the observed covariances between items by creating constructs that are the reason for the covariance. 84.1% of the variation could be explained by these 3 constructs that are identified from the EFA model. For the improved model fit, covariances were added as per the modification indices. The chi-square test is the most commonly used global fit index in CFA which has been used to test whether the covariance matrix derived from the model represents the population covariance. A number of model fit indices (chi-square value, GFI, RMSEA, AGFI etc) were determined and compared with its acceptance level to find if indicators are adequately measuring their intended concept. The lack of Fit at the first run indicated the model is a poor model of the data due to poor choice of variables. It also gave an indication that poor fit might be due to collinearity, and inclusion of irrelevant variables or exclusion of important variables. The model improved with the goodness of fit values by deleting items having poor factor loading (less than 0.5) and applying proposed modification indices. overall "fit" indicates how well the specified model is able to reproduce the original correlation analysis matrix. In nutshell, the study used factor analysis methods (EFA and CFA) which established the construct validity of the staff survey questionnaire after reducing a number of items based on the variability and observed correlation between the independent variables, hence the form refined with relevant variables and constructs in assessing the governance performance and the framework adopted in the Data and statistics department of EHS.

## **CONCLUSION:**

The results of this study provide evidence to support a three-factor (process, people, technology) representation of the data governance performance, as well as the originally hypothesized two-factor structure (process and people). Fit indices for the three-factor model were statistically superior when compared to the one-factor model and marginally better in comparison to the two-factor model.

The future scope of this study is to assess the construct validity of the data governance questionnaire by demonstrating factorial invariance using a range of demographic variables. In the further study, this three-factor structure will be verified to be invariant across the demographic groups. At this stage, a measurement model invariance test will be performed to determine if the factor loadings of the items in CFA do not differ across demographic groups. It is to prove if lack of measurement invariance is found which indicates the meaning of the unobservable construct / latent variable is shifting over time or across the groups. The findings of the current study will help to provide regulators with confidence in the properties of this instrument when considering its implementation in other units of the entity for an effective governance framework.

## **REFERENCES:**

1. Wipulanusat, W., Kokkaew, N., Parnphumeesup, P., & Sunkpho, J. (2019). Exploratory and confirmatory factor analysis of measurement scale for psychological attachment to an organization. *WMS Journal of Management*, 8(3), 1–11. Retrieved from <https://so06.tci-thaijo.org/index.php/wms/article/view/206311>
2. Marsh HW, Morin AJ, Parker PD, Kaur G. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu Rev Clin Psychol*. 2014;10:85-110. doi: 10.1146/annurev-clinpsy-032813-153700. Epub 2013 Dec 2. PMID: 24313568.

3. C. Wang, "Confirmatory Factor Analysis of Technological Innovation Dynamic Capability," *2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM)*, 2021, pp. 674-680, doi: 10.1109/ICEKIM52309.2021.00153.
4. González-Blanch C, Medrano LA, Muñoz-Navarro R, Ruíz-Rodríguez P, Moriana JA, Limonero JT, Schmitz F, Cano-Vindel A; PsicAP Research Group. Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS One*. 2018 Feb 23;13(2):e0193356. doi: 10.1371/journal.pone.0193356. PMID: 29474410; PMCID: PMC5825085.