

DOMINANT PRACTICES AND GUIDE FOR CREATING EFFECTIVE SURVEY SCALES IN SOCIAL SCIENCES

Faryal Razzaq*

Managing Director, Center for Ethical Leadership, Karachi School of Business & Leadership (KSBL), Karachi, Pakistan.
CEO The FEEEL (PVT) LTD
Faryalrazzaq11@gmail.com

ABSTRACT

The data is useless if the measure is not reliable and valid. The current paper is an attempt to bridge the gap of developmental issues regarding theory and practice of valid scales. The researches on the scale development practices found very few sound practices among management sciences practitioners and researchers engaged in new scale development (Razzaq & Aftab, 2015; Hinkin, 1995). Similarly, Fornaciari, Sherlock, Ritchie, & Lund (2005) reported in their analysis of methodologies and practices of scale development from 1994- 2004 that only 6% studied conducted confirmatory factor analysis for scale development and only 3% mentioned discriminant validity. One of the main reasons is that developing valid and accurate scales is an arduous and lengthy process, therefore, in general researchers tend to develop measures without going through the rigorous procedure of scale development (Schmitt, 1991). The trend of developing scale is escalated but the lack of rigorous procedure to develop and validate the scale remains a grey area (Razzaq, 2017). A step wise guide is proposed integrating main theories from Domain of the content to evaluating the items (Item correlations, Cronbach Alpha, EFA (PCA) reconfirmed by CFA for new scales and establishing content, Construct and Criterion validity.

Keywords: Scale Development, Validity Analysis, Questionnaire Development.

INTRODUCTION

The proposed theories are established with empirical evidence. For these scales are developed to empirically check the relationship and apply statistical procedures to confirm the hypothesized relationships. The surveys are conducted to provide quantitative descriptions to trends, opinions, attitudes by studying the sample and inferring on the population (Creswell, 2014). The goal of the survey is to derive comparable data through scales across subsets of the chosen sample so that the similarities and the differences can be found (Cooper & Schindler, 2006).

There are three ways that the management data can be obtained, by using already measured data (secondary data), or by using already established scale to collect primary data and lastly if none of the above categories provide information on the required variable then the data is collected using specially designed instruments (Krishnaswamy, Sinakumar, & Mathirajan, 2009). The measure is defined as, “the assignement of numbers to objects to present amount of degrees of a property possessed by all of the objects” (Kerlinger, 1973). Therefore, the numbers represent the same relevant relations as the objects they represent. In social sciences “scaling” means to quantify the subjective abstract measure like behaviours, so a number is assigned to some of the characteristics of the construct in question (Edward, 1957).

Hinkins (1995) in his research for scale development practices in management sciences asserted that as management sciences research is almost dependant on data collected from surveys, hence a lot of decision making is dependent on the survey. Therefore, the methodological or reporting errors in the developmental procedure could impede the progress of the discipline itself. His research found very few sound practices among management sciences new scale development and focused on three key areas of questionnaire development, Instrument item generation, Scale development and Scale evaluation.

Fornaciari, Sherlock, Ritchie, & Lund (2005) examined the new scale development studies published during 1996-2004 and analysed their methodology for scale development. Following are the main results of their study.

* Corresponding Author

- For item generation only 71 percent of the authors reported how scale items were developed. 61 percent used deductive methods from established theories for item generation. About 87 percent of studies elaborated the sample however, 77 percent clearly described their sampling technique.
- For Scale Development only 55 percent of the studies reported using factor analysis for construct evaluation. About half of the studies reporting using factor analysis did not provide factor retention criteria. Only 19 percent studies explained Eigenvalues and 26 percent explained variance of the factors. Only 6 percent of the studies reported performing confirmatory factor analysis. For internal consistency Coefficient alpha was reported in 52 percent of the sample.
- For Scale Evaluation the most surprising finding was only 32 percent reported on practices used in the area of scale evaluation. Only 3 percent mentioned discriminant validity and 29 percent tested measures on additional groups.

Although the number of empirical articles published in the MSR (Management, Spirituality and religion) domain reveals that the number of articles published between 2001-2004 was almost doubled than the scale development articles published during 1996-2000, still many researchers ignored or did not report crucial elements of new scale development, calling for more rigorous and sound measures development practices (Fornaciari, Sherlock, Ritchie, & Lund, 2005). The researchers conducted on scale development practices among management sciences revealed a lot of ambiguous methodology, entailing to study further the scale development process and its crucial elements (Razzaq, 2017; Razzaq & Aftab, 2015). Following section will discuss and converge the suggestions of different theorists for how to develop scales.

Steps for New Scale Development

Scales are mostly developed by using factor analysis (Devellis, 2012). It is done to ensure that the questions asked are relevant to the construct one desires to gauge (Fields, 2005) and explains which linear components exist within the data. Devellis (2012) in his book, "Scale development Theory and Applications" elicit following steps for new scale development in Applied Social Research Methods Series.

1. To determine precisely what is to be measured. This is done by examining the existing theories' critique and then formulating one's own theoretical framework on the gaps found in the literature. First ascertain the Domain of the content. In the next step definitions are generated and operationalized. The items are phrased according to the operationalized definitions.
2. Generate an item of pool. This is done after operational definitions are generated and based on these definitions sufficient items are generated for each variable about 3-4 times larger than the final score. However, for large scale it may be 50% more than the final scale. The final items as highlighted by the practitioners and determining the format for scale e.g. Likert scale
3. Let the initial item pool be reviewed by experts in the field. Based on their opinion and recommendations final items are retained for the developed scale. Devellis also recommends for more than 1 item per variable at least.
4. Administer the refined scale to the target population sample. The variables themselves are explained in the survey instrument.
5. Evaluate the items using item correlations, Cronbach Alpha, Exploratory Factor Analysis (Principal Component Analysis) and the items dropped in PCA are reconfirmed by Confirmatory Factor Analysis. Same data for new scale development is recommended for EFA and CFA (Devellis, 2012).

Scales are assessed for reliability as a pre-requisite of their validities and because we need stable measures free from random errors, for traits like personality and emotional Intelligence (EQ) are assumed to be stable measure and are not supposed to fluctuate in erratic ways for normal persons (Mitchell & Jolley, 2013). The questions/items in the scale not only have to be consistent but their answer should provide relevant information regarding the construct, in another words the items asked should tap the construct. Therefore, quality of the results from the scale items reflects appropriateness and accuracy of the procedures applied to answer the research questions regarding a construct, this concept is known as validity (Kumar, 2011). There are two types of validities: internal and external validity. The external validity is concerned with the generalizability of the findings of the measure outside the controlled conditions or for samples with different characteristics than the original (usually

for experimental designs), whereas internal validity is the ability of the instrument to measure the construct it was supposed to measure (Krishnaswamy, Sinakumar, & Mathirajan, 2009). There are three types of Internal validities, *Content validity* also known as logical validity, *Criterion related validity* which is further divided into predictive validity and concurrent validity and *Construct validity* which is divided into convergent and discriminant validity (Mark & Jolley, 2013; Sekaran & Bougie, 201; Zikmund, 2006).

Reliability analysis can be done through Cronbach Alpha. For validity analysis Subject Matter Experts (SME) opinion can be taken for content validity. Content validity is used in order to select the specific range of the contents (Nunnally & Bernstein, 1994). Content validity can also be inferred by the high factor loadings in EFA. However statistical significance of construct validity using EFA is established with communalities extraction equal to or greater than 0.5 (Yong & Pearce, 2013). Factors may be retained using on Eigen value rule of > 1 (Thompson & Daniel, 1996) and by the results in the rotated component matrix. The high factor loadings also show the content validity of the construct items (Fararah & Al-Swidi, 2013; Hair et.a., 2010; Chin, 1998).

To establish Criteriaian related validity it is done through establishing *concurrent validity* which is when a known population sample differ on the criteria the scale was supposed to measure, like a happiness scale should differentiate happy or unhappy people, reserachers have used criterias like performance, tenure oe even income (Dulewics & Goleman, 2000). Similarly criterion related validity can be established through *predictive validty* which is concerned with how well a future event can be forecasted by a measure for exapmle a measure that can accurately predict suitability of a candidate for a profession (Adam, Khan, & Raeside, 2014; Krishnaswamy, Sinakumar, & Mathirajan, 2009; Zikmund, 2006).

Whereas, Construct validity is concerned about a theoretical relationship between variables. It is a more sophisticated statistical procedures to establish validity of the instrument as it can only be assessed indirectly by an observable indicator as most latent constructs like EQ can not be observed directly (Adam, Khan, & Raeside, 2014; Devellis, 2012). There are two ways to establish construct validity though convergent and discriminant validity. *Convergent validity* can be established though Confirmatory Factor Analysis (CFA) though item reliability, composite reliability and average variance extracted (Fararah & Al-Swidi, 2013;Hair et.al.,2010; Bagozzi & Yi,1998) or through correlation analysis with another establish scale for the same construct (Sekaran & Bougie, 2011). *Discriminant validity* can be established either though correlation matrix of factors of the construct and putting square root of average variance extracted in diagonal places, if the diagonal score is higher the validity is established (Fararah & Al-Swidi, 2013; Fornell & Larcker,1981) or/and when the measure show nill to very weak correlations with measures of another construct like EQ should not show high crrelations with mood scales. Criterion related validity and Construct validity should be ensured for new scales (Devellis, 2012).

The research community in psychometrics have a consensus that the best way to develop questionnaires are mostly using factor analysis (Creswell, 2014; Devellis, 2012; Dyer, 2012; Bar-On ,1997a; Cooper & Schindler, 2006; Field, 2005; Stys & Brown, 2004; Mayer, Salovey, & Caruso, 2004; Boyatzis, Petrides & Furnham, 2003; Goleman, & Rhee, 1999; Jr. Raymond & Little, 2003; Dulewicz & Higgs,1999; Lewis-Beck, 1994. Factor analysis is used to measure the construct validity of the questionnaire. The validity test most commonly used is Exploratory Factor Analysis (EFA). It works by summarizing a large set of items into smaller set of components that measuring the same thing (Lewis-Beck, 1994; Brown, 2006). Therefore, it reduces larger number of dimensions into more meaningful number that enables researcher to cluster the same items together. Principal Component Analysis (PCA) transforms a number of correlated variables into a smaller number of uncorrelated variables (factors) and show how a particular variable might contribute to that component. This statistical technique clusters a number of correlating variables to factors (Lewis-Beck, 1994; Brown, 2006) and helps reduce the dimensions. The items that best describe content domain are selected.

Suitability is ensured using the Kaiser-Mayer-Olkin (KMO) test. The KMO test measures the adequacy of a sample in terms of the distribution of values for the execution of factor analysis (George & Mallery, 1999; Field, 2009). The other Bartlett's test measures the null hypothesis that initial correlation matrix is an identity matrix (Field, 2009).

Each principal component is also called as an Eigenvector. The enormity of the vector is indicated by the Eigen value which is the variance on the new factors that are successively extracted.

In factor analysis, the most common technique used for taking decision is an eigenvalue above one considered as statistically significant (Olson, Slater, & Hult, 2005) another key parameter observed is the correlation of the variable and the factors extracted which are called factor loadings. Factor loadings are important to judge which variables are related to which factors.

Factor Rotation is employed to improve the explanatory power of the variables by extracting distinct factors. The form of rotation employed in most studies is varimax which is the most common form of rotation. This is an orthogonal form of rotation which is typically used with Principal Component Analysis (Tabachnik and Fidell, 200) using Kaiser Normalization. Finally a confirmatory factor analysis and average variance extracted can be conducted to assess the construct validity.

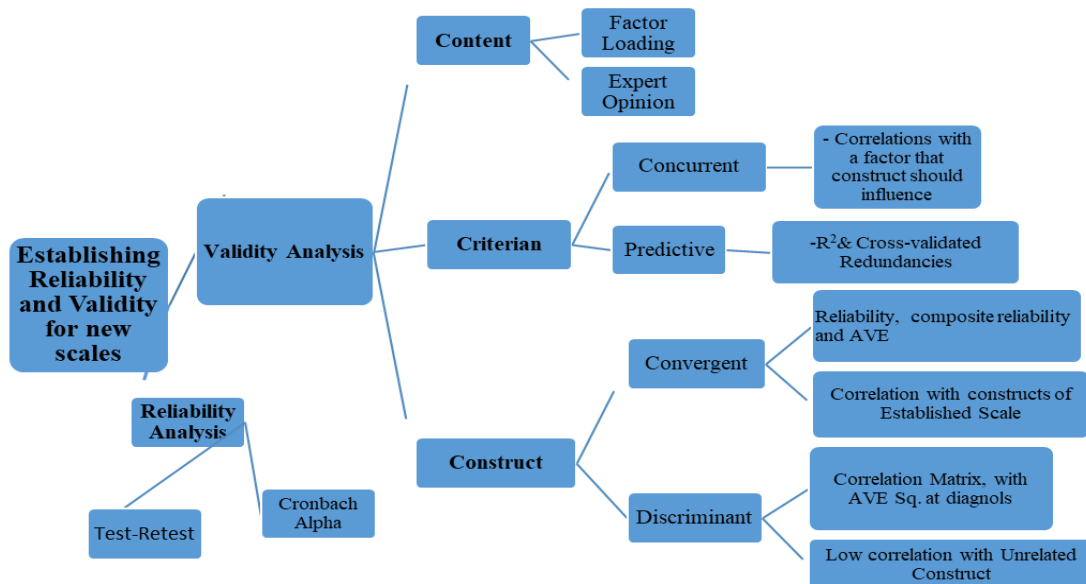


Fig. 1 Reliability and Validity Analysis Road Map for New Scale Development. Source: Razzaq (2017)

There are certain similarities between EFA and CFA as these techniques are based on assumptions of linear statistical models and assume a normal distribution. Both mentioned techniques make use of measurable variables and work on latent constructs (Holtzman, 2014). But the main difference lies where CFA requires specification of a model a priori including number of factors and which items belong to what factor, therefore it is based either on a model based on established theory or grounded in theory with errors explicitly mentioned. Whereas EFA determines the factor structure of the model by explaining maximum amount of variance (Prudon, 2015).

To sum it up following procedure is suggested to develop new scales as suggested by Razzaq (2017), Devellis (2012), Dyer (2012) Duas & Ashkanasy (2005);

1. To determine precisely what is to be measured
2. Domain of the content should be established existing literature and theories. Construct Development is an inductive process grounded in existing theories (Cronbach & Meehl, 1955)
3. Definitions are generated & operationalized. The items are phrased according to definitions.
4. Generate an item of pool (1:5 or for more than 30 variables Twice as many)
5. Determining the format for scale (e.g Likert Scale 1-5)
6. Let initial item pool be reviewed by experts in field
7. Asses for face validity and then Pilot test
8. Refine the scale
9. Administer the refined scale to the target population sample
10. Evaluate the items (Item correlations, Cronbach Alpha, EFA (PCA) reconfirmed by CFA for new scales)
11. Establish content, Construct and Criterion validity

The following figure provide a pictorial view how to establish reliability and validity for the newly developed scale the startegy employeed to establish relaiabilty and validity of EQFS© and the subsequent sections presents the results of establishing validities.

This paper was a step forward to reiterate the importance of sound practices in developing scales. Hopefully the research attention towards the field will bring theorists together to come up with novel ways to develop valid and reliable scales.

REFERENCES

- Cooper, D. R., & Schindler, S. P. (2006). *Business Research Methods* (9th ed.). NY: McGraw Hill-Irwin.
- Costello, A., & Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation, 10*(9), 1-9.
- Creswell, J. W. (2014). *Reserach Design* (3rd ed.). New Delhi: SAGE Publications India Pvt Ltd.
- Devellis, R. F. (2012). *Scale development Theory and Application* (3 ed.). (L. Bickman, & D. J. Rog, Eds.) Los Angeles, USA: Sage.
- Dulewics, V., & Goleman, D. (2000). Emotional intelligence – A review and evaluation study. *Journal of Managerial Psychology, 15*, 341-372.
- Fararah, F. S., & Al-Swidi, A. K. (2013). The Role of the Perceived Benefits on the Relationship between Service Quality and Customer Satisfaction: A Study on the Islamic Microfinance and SMEs in Yemen Using PLS Approach. *Asian Social Science, 9*(10). doi:10.5539/ass.v9n10p18
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.
- Fornaciari, C. J., Sherlock, J. J., Ritchie, W. J., & Lund, K. (2005). Scale development practices within the msr domain: an assessment of the literature from 1996-2004. *Academy of Management Conference 2005*.
- Fowler, F. J. (2014). *Survey Research Methods* (5 ed.). Thousand Oaks, California: Sage Publications, Inc.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1995). *Multivariate data analysis*. (4th ed.). New Jersey: Prentice-Hall Inc.
- Henson, R., & Roberts, J. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement, 66*(3), 6.
- Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of manegment, 967-988*.
- Holtzman, S. (2014). Confirmatory Factor Analysis and Structural Equation Modeling of Noncognitive Assessments using PROC CALIS . *SAS Global Forum 2014*. Washington, DC.
- Krishnaswamy, K., Sinakumar, A. I., & Mathirajan, M. (2009). *Management Research Methodology* (3 ed.). New Delhi: Pearson Education.
- Kumar, R. (2011). *Research Methodology: A step by step guide for beginners*. New Delhi: SAGE Publications India Pvt Ltd.
- Mitchell, M., & Jolley, J. (2013). *Reserach Design Explained* (4th ed.). Singapore: Thompson Learning. doi:0-15-507505-5
- Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: a critique. *Comprehensive Psychology, 4*(10).
- Razzaq, F. (2016). Emotional Intelligence at Work Place: A Case of Hyper Markets of Pakistan. *PhD Thesis*. Bahria, University, Islamabad.
- Razzaq, F. (2017). Emotional Intelligence at Workplace: A Case of Hypermarkets of Pakistan. Bahria University, Islamabad.
- Razzaq, F., & Aftab, F. (2015, Jan-Jun). Developing an Indigenous Scale of Emotional Intelligence. *Journal of Bussiness and Economics, 7*(1).
- Razzaq, F., Zadeh, Z. F., & Aftab, F. (2016). A theoretical evaluative study of four influential models of emotional intelligence. *Bahria Journal of Professional Psychology, 15*(2), 39-62.
- Sekaran, U., & Bougie, R. (2011). *Research Method for Bussiness: A skill building approach* (5th ed.). New Delhi: John Wiley and Sons, Ltd.

- Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems, 13*, 380-427.
- Suhr, D. (2012). Exploratory Factor Analysis with the World Values Survey. *SAS Global Forum 2012*, (pp. Paper 331-).
- Suhr, D., & Shay, M. (2009 September 1-3). Guidelines for Reliability, Confirmatory and Exploratory Factor Analysis. *SAS Conference Proceedings: Western Users of SAS Software 2009*. San Jose, California.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thompson, B., & Daniel, L. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*(2), 197-208.
- Williams, B., Brown, T., & Onsman, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3). Retrieved from <http://ro.ecu.edu.au/jephc/vol8/iss3/1>
- Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.