FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

Factor Analysis: Dealing with Response Bias

Robert Goedegebuure[1*] and Manorama Adhikari [2]
[1]*Professor and Research Director at Swiss School of Management Research Center*
[2]*Deputy Chief of Party in Monitoring, Evaluation, and Learning Project of USAID in Nepal*

***Correspondence:** Robert Goedegebuure; Email; robert.goedegebuure@ssmresearch.com

ABSTRACT- This paper proposes an innovative method for factor analyzing data that potentially contains individual response bias. Past methods include the use of "ipsative" data, or, related to that, "ipsatized" data. Unfortunately, factor analysis as the main method used for analyzing the dimensionality of data, cannot be applied to ipsative data. In contrast, normalization of data as an alternative method to filter out response bias, is not hampered by the technical statistical issues inherent to applying multivariate techniques to ipsative data. Using high-quality data from a survey in Nepal that makes use of – among others – the High-Performance Organizations (HPO) framework, this paper shows that the traditional approach of directly applying Confirmatory Factor Analysis (CFA) starting from an existing model or theory, is inferior to our approach. Even applying Exploratory Factor Analysis (EFA) to the raw (non-normalized) data before using CFA, is unable to detect the optimal dimensionality, or structure, in the data. A better structure can be obtained by performing EFA on normalized data that corrects for response bias in the raw data. This paper convincingly shows that the newly identified structure is superior to the original structure suggested by the HPO framework. Applying a CFA using the newly detected structure on the raw data, gives excellent goodness-of-fit statistics, with more items retained, and no need of forced methods to improve the model fit. The findings suggest that existing models and questionnaires based on these models, are not necessarily as valid and reliable as empirical studies that make use of traditional analyses seem to suggest. When adopting existing instruments, researchers are advised to critically check the validity and reliability of these instruments – especially those vulnerable to response bias - and to apply the procedures laid out in this paper, in order to enhance the quality of their research, and to inform future researchers who consider using the same instruments or to warn them about the potential shortcomings of these instruments.

Keywords: Exploratory Factor Analysis; Confirmatory Factor Analysis; Ipsative Data; Ipsatized Data; Normalized Data; Response Bias; High-Performance Organizations.

ARTICLE INFORMATION
Author(s): Robert Goedegebuure and Manorama Adhikari;
Received: 26/07/2022; **Accepted**: 07/01/2023; **Published**: 20/02/2023;
e-ISSN: 2347-4696;
Paper Id: IJBMR 2607-01
Citation: doi.org/10.37391/IJBMR.110103
Webpage-link:
www.ijbmr.forexjournal.co.in/archive/volume-11/ijbmr-110103.html

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

It comes as no surprise that professional and academic interest in the antecedents of organizational performance has led to a vast number of models and theories, and related questionnaires that researchers can use in their specific research setting. Research strategies range from the use of one single model and questionnaire, to the development of one's own questionnaire in the belief that no existing models capture the idiosyncrasies of the object of study. Most researchers, for good or bad reasons, revert to the use of existing questionnaires. Advantages of the latter strategy are the relative ease in assessing the reliability and validity of the measurement instrument, and the ability to compare one's results to those of previous similar studies. A disadvantage of adopting existing questionnaires is the loss of flexibility, reflected by the inclusion of items that are irrelevant to the case at hand and the exclusion of potentially relevant items (Goedegebuure, 2018). An in-between strategy that combines the advantages of using existing questionnaires and the development of one's own questions, would often be better. Whichever strategy is used, researchers tend to take the relevance, validity and reliability of items taken from existing questionnaires for granted.

This paper is based on research on the performance of three Nepalese public sector organizations. The research has made use of a mixed strategy. The basis of the measurement instrument was De Waal's High-Performance Organization (HPO) framework and the related questionnaire consisting of 35 items which are assumed to belong to five factors (continuous improvement; openness & action orientation; management quality; employee quality; and long-term orientation; De Waal, 2007). Items were added based on literature review and in-depth interviews with employees and managers of the three organizations. This paper critically reviews the HPO-part of the resulting data (n=300; 100 respondents per organization examined).

One particular worry is that all 35 items and the five factors implied by De Waal are highly intercorrelated. The factors and retained items of the HPO model have been derived from an exploratory factor analysis of many items that were found in HPO-related literature. The resulting five-factor structure seemingly have been confirmed in subsequent studies that have used the HPO questionnaire as the sole basis (e.g., De Waal, 2010; De Waal et al, 2009, 2013).

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

Factor analysis, either exploratory or confirmatory, is based on a matrix of correlations between the items. High correlations between items can result from (*i*) the fact that items indeed belong to one factor; (*ii*) individual response bias; and (*iii*) a halo-effect.

Individual response bias implies that some respondents tend to give lower or higher scores than other respondents, even though their true opinions do not differ.

A halo-effect occurs when one's general good or bad feeling about a topic (in this case, the organization) causes scores on any issue related to the organization to be higher or lower. If these two biases make up a substantial part of the data-generating process, then the correlations between items become inflated, and important information on the true structure in the data is obscured.

We can illustrate this with a stylized example.

	i1	i2	i3
1	1	1	3
2	1	2	2
3	1	3	1
4	3	4	2
5	3	3	3
6	3	2	4
7	7	6	8
8	7	7	7
9	7	8	6
10	8	7	9
11	8	8	8
12	8	9	7
13	9	10	7
14	9	9	9
15	9	8	10

Figure 1: Stylized Sample Data

The example of *figure 1* shows the scores of 15 respondents on three items, **i1** to **i3**. The items are normatively measured, using a 10-point Likert Scale. Let's assume that item **i1** can be regarded as the general feeling about a topic. At any level of **i1**, the values of **i2** and **i3** are set in such a way that if one increases, the other decreases. The implied negative correlation between **i2** and **i3** is obscured by the fact that across respondents all scores move in the same upward direction. Without correcting for response bias, the correlation between items **i2** and **i3** is 0.82, which is generally considered as strongly positive.

Ipsative measurement or forced-choice techniques, as an alternative to normative measurement, have been in use since the 1950s. Respondents are forced to choose between options of equal desirability that are more or less true to them. The most (least) true option chosen then earns most (fewest) points. Such scores reflect intrapersonal comparisons.

If our scales are not ipsative but, like in *figure 1*, normative, then we can still filter out individual biases (individual response bias plus halo-effect) by *ipsatizing* the data. In this procedure, the scores for each respondent are rescaled with the individual mean as a point of reference. For example, one can use a Z-transformation for the item-scores for each respondent, to compute *ipsatized* scores. The Z-scores are computed as:

$$I_{n,ips} = \frac{(I_n - MEAN_{I.})}{SD_{I.}}$$

In this formula, for any respondent $I_{n,ips}$ is the ipsatized score on the n^{th} item, out of *N* items; I_n is the (raw) score on the n^{th} item; $MEAN_{i.}$ is the mean of the raw scores on the *N* items; and $SD_{i.}$ is the standard deviation of the raw scores on the *N* items. The ipsatized scores are given in the *figure* below. The correlation between items **i2** and **i3** is now highly negative (-.92), which is in line with the way the data were generated. Note that ipsatized scores using the Z-transformation cannot be computed if the standard deviation of the item scores equals zero.

	i1	i2	i3	meani	sdi	i1s	i2s	i3s
1	1	1	3	1.67	1.15	-0.58	-0.58	1.15
2	1	2	2	1.67	0.58	-1.15	0.58	0.58
3	1	3	1	1.67	1.15	-0.58	1.15	-0.58
4	3	4	2	3.00	1.00	0.00	1.00	-1.00
5	3	3	3	3.00	0.00	.	.	.
6	3	2	4	3.00	1.00	0.00	-1.00	1.00
7	7	6	8	7.00	1.00	0.00	-1.00	1.00
8	7	7	7	7.00	0.00	.	.	.
9	7	8	6	7.00	1.00	0.00	1.00	-1.00
10	8	7	9	8.00	1.00	0.00	-1.00	1.00
11	8	8	8	8.00	0.00	.	.	.
12	8	9	7	8.00	1.00	0.00	1.00	-1.00
13	9	10	7	8.67	1.53	0.22	0.87	-1.09
14	9	9	9	9.00	0.00	.	.	.
15	9	8	10	9.00	1.00	0.00	-1.00	1.00

Figure 2: Stylized Sample Data with Ipsatized Scores

A related issue is the impact of response bias on commonly used statistics that justify the use of factor analysis: the Bartlett test and the KMO score. While the Bartlett test of the correlation matrix being an identity matrix only in rare cases would suggest against the use of factor analysis in the practice of social and economic studies, the KMO test makes intuitive sense and is easy to interpret as the statistic comes with labels, ranging from *miserable* for values of around 0.50, to *marvelous* for values close to 1.00 (Cerny & Kaiser, 1977).

As a benchmark, a data set was generated with the same size as the actual data set (n=300) and a data-generating process with a uniformly distributed individual *"HPO level"* for all 300 cases, and a random normally distributed deviation (mean=0; standard deviation=1.2) from the individual level, for all 35 items. Factor analysis of the random data set suggests, as expected, exactly one factor. The KMO statistic is 0.99, which would read as "marvelous" to any naïve researcher.

In the left-hand panel of *figure 3*, the scree plot for the actual data is shown, and it is obvious that it is not much different from the randomly generated benchmark, with a "marvelous" KMO-statistic of 0.95. According to commonly applied rules-of-thumb, like visual inspection of the "elbows" in the scree plots or keeping factors with eigenvalues exceeding 1 (Field et al, 2012), two factors would be retained. On the basis of parallel analysis, considered to be the superior way to determine the

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

number of factors to be retained (Ledesma & Valero-Mora, 2007), as many as nine factors would be kept. Interestingly, when – for whatever reason - deciding on five factors (as in De Waal's framework; De Waal, 2007), the "theoretical" structure of factors and their items as suggested by De Waal's model, is, by and large, obtained.

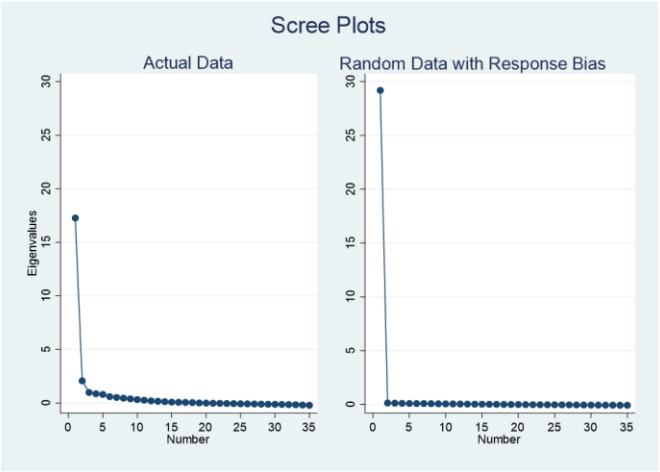

Figure 3: Scree Plots of Actual Data versus Random Data with Uniformly Distributed Individual Levels

The research question in this paper is to see if the five-factor structure proposed by De Waal is confirmed by in-depth analysis of the data.

More in particular, this paper will use both exploratory factor analysis and confirmatory factor analysis (EFA and CFA, for short) to find the structure in the data assuming the raw (normative; non-ipsatized) data may suffer from response bias.

After a review of the literature on ipsatizing data, this paper will suggest a procedure that effectively deals with the assumed response bias in the data. More in particular, EFA will be applied to normalized data. In order to illustrate the added value of this approach, this new method will be contrasted to the common approach of directly applying CFA on the raw data.

The rest of this paper is organized as follows. In the next section we will review the literature on the use of ipsative or ipsatized data in factor analysis. With the lessons learned from the review of literature, section 3 will proceed with factor analyzing the raw data and the normalized data, and make a comparison of the outcomes. The final section will draw conclusions.

2. LITERATURE REVIEW

Response bias is a challenge faced by researchers when using Response bias is a challenge faced by researchers when sampling scaled data from different populations. For example, different cultures are known to have different types of responses, due to seeking or avoiding of the extremes of the scale, which hampers comparisons across cultures or groups (Cunningham, 1977). Already in the 1950s, ipsatizing data was seen as a method with great potential for reducing response bias. The term ipsative is derived from the Latin *ipse*, which means "himself". Ipsative or ipsatized data can be defined as any score matrix characterized by a constant sum of scores across respondents. The notion of ipsative scales was introduced by Catell (1944), who distinguished ipsative scales from interactive scales and normative scales. Ipsative scales measure attributes of individuals relative to other attributes of the same individual.

Chan (2003) discusses various types of ipsative data, namely additive, multiplicative and ordinal ipsative data. Some commonly used ipsative methods are constant sum scales; data on percentages of household expenditures by category; or row-centered or row-standardized scores (Chan, 2003; Hofstee et al., 1998).

Applications and justifications for ipsatized data can be found in many studies (Moskowitz, 2005; Chan, 2003). Several researchers over time (Hicks, 1970; Cunningham, 1977; Chan, 2003) have warned that there has to be a clear justification for using ipsatizing data. One justification is the presence of a systematic response bias. In analogy to the logic of using ipsative data stemming from different cultures, ipsative data can be justified in case respondents are employees from one and the same organization. A systematic difference in the level of scores across respondents on items that are part of – at least theoretically - unrelated factors (like for example employee quality and long-term orientation in the HPO-model), suggests individual response bias. Such score patterns are likely to reflect general feelings about the topic as a whole (e.g., the organization) rather valid assessments of the dimensions of the topic. That is, the data generating process is likely to resemble the data generating process used in *figure 1*.

One important source of response bias is acquiescence. Acquiescence is an overall tendency to endorse items. To the extent that respondents show the same level of acquiescence it adds a constant that leaves the correlational structure intact. Differential acquiescence due to some respondents agreeing more to items than other respondents, inevitably increases the correlation between items (Hofstee et al., 1998). The same holds true for socially desirable answers: differences between respondents in the impact of social desirability on their scores, affects the observed correlations between items. While acquiescence and social desirability are assumed to happen unconsciously, response faking is a deliberate act by the respondent to overreport or underreport his or her true feelings. Seeking and avoiding of extremes is a closely related phenomenon.

While ipsatizing data to correct for individual response bias makes intuitive sense, many researchers (Chan & Bentler, 1993; Closs, 1996; Johnson et al., 1988; Cornwell & Dunlap, 1994; Moeller, 2015) argue that correlational methods like factor analysis cannot be applied to ipsative data. Ipsative and ipsatized data produce matrices that fail to reflect the true correlations between items. A requirement for methods based on correlations, is that the variables are statistically independent, which is simply not the case for ipsative or ipsatized data (Closs, 1976, 1996; Van Eijnatten et al., 2015). This is not to deny that ipsative methods are useful if one is

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

interested in the relative ordering of items within a person, and even across persons (Cornwell & Dunlap, 1994). However, as Closs (1996) has it, ipsative or ipsatized information indicates what is preferred rather than what is liked. Absolute rather than relative measurement can only be obtained by using non-ipsative methods.

Among the few proponents of factoring ipsative or ipsatized data, are Saville and Willson (1991) who have used simulated and randomized data to come to the conclusion that ipsative data can be legitimately factorized. Chan (2003), while acknowledging the problems when applying factor analysis to ipsative data, suggests some advanced methods for extracting information from ipsative data sets.

Closs' (1996) findings that factorizing ipsative matrices gives rise to spurious bipolar factors, is confirmed in our own trials with both empirical data and randomly generated data sets based on given factor structures. Still, it has to be said that data sets with response biases in normative (non-ipsative) data too generate spurious results. While Closs (1996: 43) provides an example of positively correlated non-ipsative items becoming negatively correlated after ipsatization, examples can be construed from random scores on items per individual in case individuals show response biases. In the latter case, spurious positive correlations will result.

In conclusion, both the literature on factor analysis of ipsative data and our own experiments using randomly generated data with individual response biases, suggest that ipsative (or ipsatized) data should not be analyzed using factor analysis. Even solutions to the technical issues like leaving out one variable from the analysis or assuming alleged robustness in case of many variables, in our view do not offer a way out. It may be that the firm statements in the literature have contributed to ipsatization - and in its wake other methods to correct for response bias - falling out of grace. According to Closs (1996: 46): "there is perhaps a need for authoritative bodies [..] to work more closely with established test publishers [..] to put a stop to it".

Rather than ipsatizing the data using a Z-transformation (leading to an average score of zero for all respondents), an alternative approach is to use *normalized* data. This approach uses intuitive logic for correcting response bias while circumventing the technical problems encountered when ipsatizing the data. This route will be explored in the next section.

3. DATA ANALYSIS

3.1 Introduction

The objective of the data analysis is to address two issues simultaneously.

First, in order to avoid taking the five-factor model that lies at the basis of the standard (HPO) questionnaire adopted in the research for granted, both exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA) have been applied to the raw data. This deviates from the common approach to use only CFA when dealing with data based on an existing model, since exploration of the data is not needed. In the same vein, CFA is not apt for the exploration of other structures than the structure implied by the underlying model.

Secondly, in order to deal with the potential consequences of response bias due to acquiescence and halo-effects in our data set, the structures resulting from EFA of both the original data and data normalized at the level of individual respondents, are compared. In order to verify if any different factor structure resulting from the EFA of normalized data performs better, CFA using both the original model and the revised model have been applied to the raw (non-normalized) data.

In order to make a fair comparison, identical criteria were applied in the EFAs of the original and the normalized data: (i) the maximum number of factors retained, is based on parallel analysis; (ii) the number of factors retained is lowered by one at the time, until all factors have at least two items with a loading>.50; (iii) items are kept if the primary loading>.50, and the second highest loading<.20 or substantially lower (.40) than the primary loading. Moreover, the varimax method has been used in rotating factors for easier interpretation rather than the oblique rotation that would allow similar factors to be differently correlated in the analyses of raw versus normalized data. For a justification of the criteria used, we refer to Hayton et al. (2004); Ledesma & Valero-Mora (2007); and Matsunaga (2010).

Figure 4 summarizes the data analysis strategy.

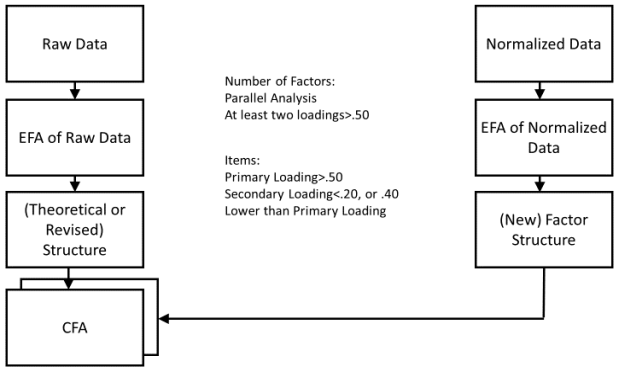

Figure 4: Data Analysis Design

Normalizing of the original scores at the individual level, uses the following formula.

$$V_{i,j}^{N} = (V_{i,j} - (\min(V_{i,.}) - 1))/((\max(V_{i,.}) + 1) - (\min(V_{i,.}) - 1))$$

In order to avoid missing values for respondents with a constant score on all 35 items, a small number (1) has been subtracted from (added to) the minimum (maximum) value for each respondent. A smaller number than 1 can be used (e.g., 0.1), but the outcomes are robust against the choice.
Table 1 gives a stylized example for 4 respondents and 3 items. The first two rows (respondents) differ in the position on the rating scale, but have the same scores after normalization. The

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

third respondent uses the full length of the scale, but – due to the stretching of the range of the scale – the normalized scores range from 0.09 to 0.91 rather than from 0 to 1.

Table 1: Applying the Normalization Formula

Original data			Min-1	Max+1	Normalized data		
v1	v2	v3			v1n	v2n	v3n
2	4	5	1	6	0.20	0.60	0.80
6	8	9	5	10	0.20	0.60	0.80
1	9	10	0	11	0.09	0.82	0.91
2	2	2	1	3	0.50	0.50	0.50

3.2 Factor Analysis of Original Data

The data set consists of 300 employees and managers at three public organizations in Nepal who have rated their organizations on De Waal's 35 items on a 10-point scale (1=very bad; 10 excellent). The data set used in the analysis contains no missing values. Respondents, during the face-to-face interviews, were allowed to use midway scores (*e.g.*, 6.5) in case they couldn't choose from the adjacent points of the scale (6 and 7). In the data set, these few occurrences were randomly set to one of the adjacent points of the scale (6 or 7). The overall impact on the correlations between items is negligible.

The (595) pairwise correlations between the 35 items are approximately normally distributed with a mean of 0.48.

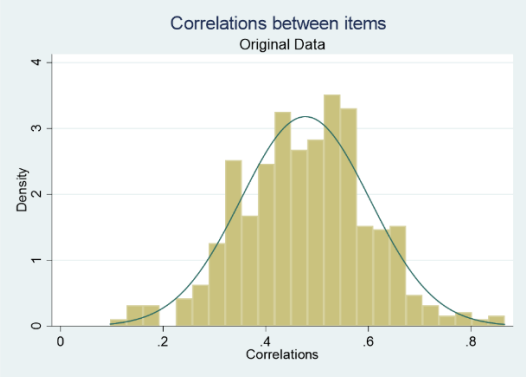

Figure 5. Distribution of Correlation Coefficients Between Item Pairs (Raw Data)

The standard criteria for factor analyzability are sufficient: the Bartlett test clearly rejects the hypothesis that the correlation matrix is an identity matrix, and the KMO-test reveals a score of 0.95, which is marvellous in the common interpretation.

The scree plot suggests that there is one dominant factor accounting for 73% of the variance in the data. Commonly used criteria for the number of factors to be retained, are *eigenvalue>1* and a visual inspection of the *elbow* in the scree plot or the levelling off of eigenvalues. Both criteria would suggest two factors. Parallel analysis, as an alleged superior method (Ledesma & Valero-Mora, 2007; Matsunaga, 2010) for deciding on the maximum number of factors to be retained, suggests up to 10 factors. Starting from 10 factors, application of the decision criteria discussed above, have brought the number of retained factors down to 6 factors with 2 or more items with a *primary loading>.50*. Items are kept if the *primary loading>.50* and the *second highest loading is <.20* or the difference to the primary loading exceeds .40 (for example, .70 and .25). As a consequence, the 6th factor has been left out.

The resulting five-factor structure resembles De Waal's model quite well. However, almost half of the items (16 out of the 35) have been dropped.

Table 2: Summary Results of Exploratory Factor Analysis of Raw Data

De Waal's HPO Model	Items (theoretical)		Retained items EFA (Nepalese data)	
Continuous Improvement (CI)	v1-v8	8 items	v2-4; v8	4 items
Openness & Action Orientation (OAO)	v9-v14	6 items	v9; v11	2 items
Management Quality (MQ)	v15-v26	12 items	v15-v19; v21-v23	8 items
Employee Quality (QEMP)	v27-v30	4 items	v29; v30	2 items
Long-Term Orientation (LTO)	v31-v35	5 items	v33-v35	3 items
Total	35 items		19 items	
Bartlett's test: KMO-test: Number of factors (eigenvalue>0): Number of factors (parallel analysis):	8,895, degrees-of-freedom 595; p=0,00 0.954 2 10			

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

3.3 Factor Analysis of Normalised Data

The correlations between the items after normalization at respondent level, are approximately normally distributed with a mean of 0.23

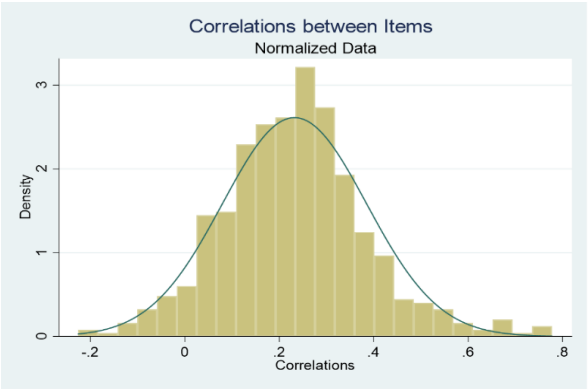

Figure 6: Correlation Coefficients between the Items (Normalized Data)

Using the exact same procedure as for the raw data, there is a less dominant first factor accounting for 51% of the variance. The scree plot levels off at four factors, while the fifth factor still has an eigenvalue just above 1. Parallel analysis indicates a maximum number of factors of 13 to be retained. Following the identical procedure (keeping factors with 2 or more items with *primary loadings>.50* and *secondary loadings<.20*), nine factors are retained. Some of the factors turn out to be splits of the factors found for the original data. Others are new combinations of items previously belonging to other factors, or previously excluded. Under the new structure, a larger number (24 instead of 19) items is retained.

Table 3: Summary Results of Exploratory Factor Analysis of Raw Data

De Waal's HPO Model	EFA of Original Data		EFA of Normalized Data		
					Tentative Label
CI	v2-4; v8	4 items	v2-v4	3 items	As in HPO
			v7-v8	2 items	
OAO	v9; v11	2 items	v9; v11	2 items	
MQ	v15-v19; v21-23	8 items	v15-v19	5 items	
			v21-v23	3 items	
QEMP	v29; v30	2 items	v29; v30	2 items	
LTO	v33-v35	3 items	-		
			v13; v14; v31	3 items	Long-term Performance
			v1; v35	2 items	Organizational Identity
			v33; v34	2 items	Organization as a Family
			Total	24 items	

3.4 Comparison Using CFA

A common strategy used when adopting existing models and questionnaires based on these models, is to use CFA to test if the data fit the model underlying these questionnaires. Under that strategy, some of the poorly fitting items may be dropped from the scales. A poor model fit can be improved via inspection of the so-called modification indices. An often used tactic for model improvement is the addition of covariances between the error terms. Although there may be a theoretical justification to do so, the final fit model may be burdened with many such terms for which theoretical justifications are not always clear. The alternative would be to drop more items, or even factors.

In the example of HPO, the direct application of CFA may end up in, for example, four factors and only 16 items. In the conclusion, the researcher would have to argue why – in the context of the specific application of the HPO-model – the theories and the scales only partially apply.

As indicated above, performing EFA before CFA on the raw data, does not lead to new insights, at least not in this case. The EFA of the raw data results in the "theoretically correct" five factors with a subset of 19 out of the 35 items from the original model and questionnaire. Most likely, directly applying CFA not preceded by EFA, would have led to very similar results in terms of items to be dropped.

In contrast, applying EFA using normalized data leads to a larger number of factors (9), and from a first inspection the new groupings make sense. In fact, from our long experience in running CFAs on data collected using the HPO-model and related questionnaire, the new groupings capture many of the frequently encountered covariances between error-terms. Whether the new structure (based on EFA of normalized data) indeed makes sense, can be tested by applying it in a CFA using the raw data.

The test is done in two steps.

- The first step is the traditional approach of applying the theoretical structure to the data. That is, the theoretical five-factor structure implied by the HPO-model is applied in a CFA to the raw data. For a fair comparison, the bigger set of 24 items retained in the EFA of the normalized data has been used. Items are assumed to be measurements of the theoretical five factors. Error-terms are not allowed to co-vary.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

- In the second step, the newly found nine-factor structure with the same 24 items is applied in a CFA to the raw data.

In both steps, goodness-of-fit statistics are computed. No attempts have been made to improve the models (no dropping of items; no adding of covariances between error terms). Leaving out the covariance terms between the latent variables (factors), and the path loadings, for readability, the two models and their goodness-of-fit statistics are depicted in *figures 7* and *8* below.

Using the 24 items retained in the EFA of the normalized data (which is a bigger set than in the EFA of the raw data) and the theoretical paths of these items to the factors implied by De Waal, a poor fit results (RMSEA is well above the strict upper limit of 0.07; the CFI is below the 0.95 criterion of a well-fitting model; and the SRMR is above the recommended 0.05 level; cf. Hooper et al, 2008).

Typically, researchers would examine the modification indices, and add paths, covariances between error terms, or drop variables or even factors in order to obtain a better fitting model. Adding paths is not an option, as from the perspective of discriminant validity, items are not allowed to be measurements of more than one factor. Dropping items and factors would make the model drift away further from the 35-item and five-factor structure that was the starting point. That leaves the researcher with the option to add covariances between the error terms of the measurements, which is often hard to justify theoretically.

In contrast, the nine-factor structure derived from the EFA of normalized data performs remarkably well, without any need to drop variables or to add covariance terms. All goodness-of-fit statistics indicate a good model fit, even though the chi-square statistic is still significantly different from zero (as would occur in a perfectly fitting model).

It turns out that a better structure in the data can be obtained by factor analyzing data after correction for response bias. Strikingly, this better structure works quite well on the raw data.

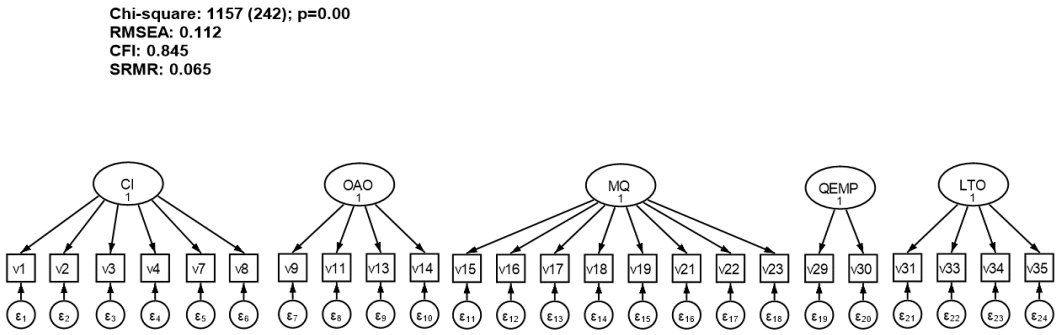

Figure 7: CFA of Raw Data using the Theoretical Factor Structure

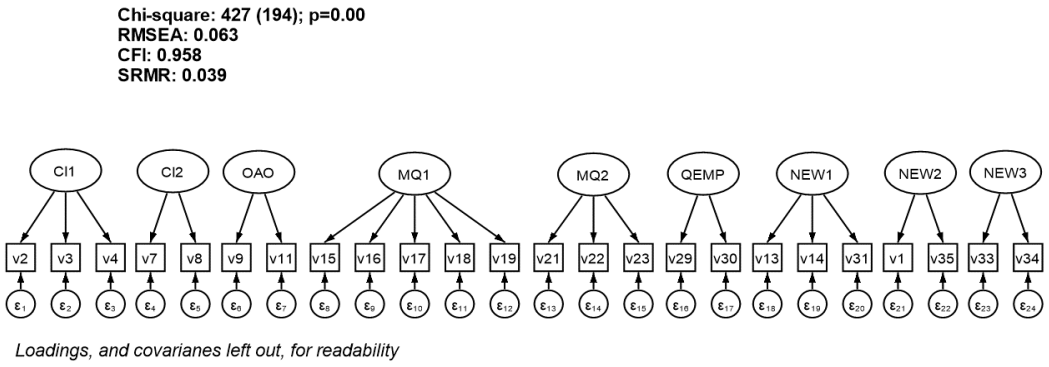

Figure 8: CFA of Raw Data using the Newly Identified Factor Structure

International Journal of Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

4. SUMMARY AND CONCLUSION

This paper simultaneously addresses various related issues in quantitative organizational and management research through surveys.

Researchers in these domains can find inspiration from a wide range of models, and related survey questionnaires. While a careful selection of one or more standard questionnaires, or parts thereof, is efficient and avoids reinventing the wheel, researchers adopting such questionnaires tend to overestimate their validity, reliability, and relevance in the context of their own research. One issue is that the validity and reliability of the measurement tools are only as sound as the often unknown or unreported assessment by the original researchers.

As an example, this paper examines the validity of a scale on high performance organizations developed and promoted by De Waal (De Waal, 2012) as a scientifically validated scale, and even the holy grail of high-performance models. The part of the data collected using the HPO model, shows a common feature in several HPO researches, namely that the 35 items belonging to five HPO-factors are strongly intercorrelated. It is conceivable that the high correlations between all items, are caused by acquiescence and a general good or bad feeling about one's organization. The challenge of dealing with response bias in statistical analyses has been addressed in literature, and led to the idea of *ipsatizing* the scores, where respondents are used as their own point of reference. Ipsatized scores provide relevant (descriptive) information about intra-individual scores, and thus about the grouping of items. Appealing as it seems, it is unfortunately not possible to factor analyze ipsatized scores; even technical solutions to statistical challenges, do not allow for substantive interpretations. A solution that comes close to the purpose of ipsatizing, and does not suffer from these challenges, is achieved by a factor analysis of normalized data – as explored in this paper. With real-life data from an organizational performance research in Nepal, it is shown that an (exploratory) factor analysis of normalized data generates a structure with more factors and more retained items than an exploratory factor analysis of the raw data. Remarkably, the new structure was not detected when applying an EFA to the raw data. It is likely that the response bias in the original data leads to one dominant factor and in the process obscures some of the delicate structures in the data. Applying the revised nine-factor model (as derived from the EFA of normalized data) in a CFA to the original data, leads to excellent results, while a CFA using the five-factor theoretical model of De Waal, leads to poor goodness-of-fit statistics; improving the theoretical model can only be achieved by adding covariances between error terms which are theoretically hard to justify, or by dropping even more items or even factors. In specific terms: even a model with only 19 items that is nested in the original 35 items framework of De Waal would perform poorly, while a revised 24 item and nine-factor model performs very well.

In general, this paper recommends researchers adopting existing standard questionnaires, to check how validity and reliability have been determined empirically in previous studies. It always pays off to apply exploratory factor analysis, using recommended standards (including parallel analysis). This paper shows that a non-conventional method of factor analyzing normalized data, reveals structures in the data that are not (yet) part of the theory.

In general, especially in organizational and management research, researchers should be critical when adopting existing questionnaires even if (or maybe especially when) they are claimed to be scientifically validated. In the specific case of De Waal's model, its application to a high-quality data set from Nepal suggests that – leaving aside the fact there are many similar instruments to pick from when hunting for questionnaires on organizational performance – it's not the holy grail as suggested by De Waal. From a positive perspective, if the HPO questionnaire has the religious miraculous powers attributed to the Holy Grail, then this paper is the first to scientifically bring us a step closer to understanding its functioning and to fine-tuning it, by analyzing its structure and by weeding out the redundant parts.

REFERENCES

[1] Baron, H. (1996). Strengths and limitations of ipsative measurement. Journal of Occupational and Organizational Psychology, 69, 49-56.

[2] Bowen, C.-C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. The International Journal of Organizational Analysis, 10, 240-259.

[3] Cerny, C.A., & Kaiser, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behavioral Research, 12(1), 43-47.

[4] Chan, W. & Bentler, P. M. (1993). Covariance structure analysis of ipsative data. Sociological Method and Research, 22, 214-247.Closs, S. J. (1996). On the factoring and interpretation of ipsative data. Journal of Occupational and Organizational Psychology, 69, 41-47

[5] Closs, S.J. (1976). Ipsative vs. Normative Interpretation of Interest Test Scores or "What Do You Mean by Like". Bulletin of the British Psychological Society, 28, 289-299

[6] Closs, S.J. (1996). On the Factoring and Interpretation of Ipsative Data. Journal of Occupational and Organizational Psychology, 69, 41-47

[7] Chan, W. (2003). Analyzing ipsative data in psychological research. Behaviormetrika, Vol. 30, No. 1, 99-121

[8] Cornwell, J.M. & Dunlap, W.P. (1994). On the Questionable soundness of Factoring Ipsative Data: A Response to Saville & Wilson. Journal of Occupational and Organizational Psychology, 67, 89-100

[9] Cunningham, W.H., Cunningham, I.C. M. & Green R.T. (1977). The Ipsative Process to Reduce Response Set Bias, The Public Opinion Quarterly. Vol. 41, No. 3 (Autumn), 379-384

[10] Davis, T.M. & Chissom, B. (1981). Factor Analysis (R-Technique) of Ipsatized Data May Be Misleading, Psychological Reports. Vol 49, Issue 2

[11] De Waal, A.A. (2007), The Characteristics of a High Performance Organization. Business Strategy Series, August

[12] De Waal, A.A., Duong, H. & Ton V. (2009), High Performance in Vietnam: The Case of the Vietnamese Banking Industry, Journal of Transnational Management, 14, 179–201

[13] De Waal, A.A. (2010), Achieving high performance in the public sector. What needs to be done? Public Performance & Management Review, 34, 1, 81–103

[14] De Waal, A.A. & Akaraborworn, C.T. (2013). Is the high performance organization framework suitable for Thai organizations?, Measuring Business Excellence, Vol. 17 Issue: 4, 76-87

International Journal of
Business and Management Research (IJBMR)
Survey Report | Volume 11, Issue 1 | Pages 25-33 | e-ISSN: 2347-4696

[15] De Waal, A.A. (2012). What Makes a High Performance Organization: Five Factors of Competitive Advantage that Apply Worldwide 2nd Edition? HPO Center

[16] Van Eijnatten, F.M., Van der Ark, L.A. & Holloway, S.S. (2015). Ipsative Measurement and the Analysis of Organizational Values: an Alternative for Data Analysis. Qual Quant, 49, 559-579

[17] Field, A., Miles, J. & Field, Z. (2012). Discovering Statistics Using R. Sage Publications.

[18] Goedegebuure, R.V. (2018). Sampling items from lengthy scales using Item Response Theory (IRT); The case of the High-Performance Organizations (HPO) Scale. StatMind Working Paper Series 2018:01. Retrievable from http://statmind.org/Publications/ (last accessed 18 February 2018)

[19] Gordon, L. V. (1951). Validities of the forced-choice and questionnaire methods of personality measurement. Journal of Applied Psychology, 35, 407-412.

[20] Hayton, J.C., Allen, D.G. & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. Organizational Research Methods, Vol. 7 No. 2, April 2004 191-205

[21] Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. Psychological Bulletin, 74, 167-184

[22] Hofstee, W.K.B., Ten Berge, J.M.F. & Hendriks, A.A.J. (1998). How to score questionnaires? Personality and Individual Differences, 25, 897-909

[23] Hooper, D., Coughlan, J. & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. The Electronic Journal of Business Research Methods, Volume 6, Issue 1, 53 – 60.

[24] Johnson, C. E., Wood R. & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: the use of ipsative personality tests. Journal of Occupational Psychology, 61, 153-162.

[25] Ledesma, R.D. & Valero-Mora, P. (2007). Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. Practical Assessment, Research & Evaluation, Volume 12, Number 2, February.

[26] Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts and How-to's. International Journal of Psychological Research, 3(1), 97-110.

[27] Dr. Diane Hamilton (2022), Developing and Testing Influencers of Perception in the Workplace with the Perception Power Index (PPI). IJBMR 8(4), 120-123. DOI: 10.37391/IJBMR. 080406. https://ijbmr.forexjournal.co.in/archive/volume-8/ijbmr-080406.html

[28] Moskowitz, D.S. (2005). Unfolding Interpersonal Behavior. Journal of Personality 73:6, December 2005.

[29] Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. Journal of Occupational Psychology, 64, 219-238.