

Piotr Tarka^{*}

LIKERT SCALE AND CHANGE IN RANGE OF RESPONSE CATEGORIES VS. THE FACTORS EXTRACTION IN EFA MODEL

Abstract. The objective of the article is the comparative analysis of Likert rating scale based on the following range of response categories, i.e. 5, 7, 9 and 11 in context of the appropriate process of factors extraction in exploratory factor analysis (EFA). The problem addressed in the article is related primarily to the methodological aspects, both in selection of the optimal number of response categories of the measured items (constituting the Likert scale) and identification of possible changes, differences or similarities associated (as a result of the impact of four types of scales) with extraction and determination the appropriate number of factors in EFA model.

Key words: Exploratory factor analysis, Likert scale, experiment research, marketing.

1. INTRODUCTION

Rating scales are among the most widely used measuring instruments in social sciences research. It is therefore not surprising that a great deal of research should be dedicated to the effects of variations in rating scale format, including differences in the number of response categories. These issues set the main aim of the investigation conducted in the present article, whose empirical results are reported below and which provides a thorough assessment of scores derived from Likert scales (varying in number of response categories) in reference to the process of factors extraction. In the course of extraction, the author described a possible impact of different scales on the performance of factors extraction. With this objective in mind, the following check criteria were applied: the item communalities, total variance explained matrices of rotated factors along with their calculated coefficient reliabilities, Kaiser-Meyer-Olkin measure of sampling adequacy as well as Bartlett's test of sphericity.

^{*} Ph.D., Department of Market Research, Poznan University of Economics.

2. GENERAL ASSUMPTIONS OF EXPLORATORY FACTOR ANALYSIS (EFA)

When assigning a place to factor analysis in the general field of statistics, we should follow Kendall, who drew a distinction between the analysis of *dependence* and *interdependence*. As he explained (Kendall 1950): „in the latter we are interested in how a group of variables are related among themselves, whereas in the former type of analysis we are interested in how a certain specified group (i.e. dependent variables) depend on the others”. Thus, the position of Exploratory Factor Analysis (EFA) in the selected group of techniques lies in the heart of the interdependence analysis.

In the construction of EFA model, researchers explain the outcome of p observable variables. EFA is but a template imposed upon the correlations among a set of variables to see what things would be like in context of the variation of these variables produced by variation in a set of common variables (Mulaik 1990). This variation is examined on the basis of EFA model as follows (Gorsuch 1974):

$$\mathbf{X} = \mathbf{\Lambda}\mathbf{F} + \mathbf{V}_{\delta_i}. \quad (1)$$

where: $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_p]^T$ – vector observed variables p , $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_k]^T$ – vector of common factors; $\mathbf{\Lambda} = [\lambda_{ij}]$, – matrix of loading factors. \mathbf{V}_{δ_i} – unique elements of every observed variable.

The most important assumptions underlying EFA are: 1) factors F_j should be centered, standardized and uncorrelated (without covariance), 2) measurement errors δ_i cannot be correlated with each other. They should have zero correlation with the common factors and zero expected value, 3) observed variables must be standardized. Moreover, since there is no covariance between F_j and δ_i or between δ_i and δ_j , then:

$$\text{Cov}(F_j, \delta_i) = \text{Cov}(\delta_i, \delta_j) = 0. \quad (2)$$

Finally, we need to stress a few distinctive functions of EFA. First and foremost, EFA is a type of factor analysis where its purpose is to identify the underlying dimensional structure, if any, of a set of items. EFA is often used in the initial stage of data exploration, e.g., in case of the multidimensional scale development. Kim and Mueller (1978) argued that the main motivation behind this special use of factor analysis is not only in ascertaining the factor structure

among a set of variables, but in achieving data reduction and obtaining better factor scales which can be used in a different studies. Besides, as Mulaik explained (1987: 268): “EFA is regarded rather as a hypothesis-generating method, (providing information for the researcher to use in formulating scientific hypotheses), than analysis depending on hypothesis-proving, as it is in confirmatory factor analysis.”

At last, in EFA, a simple structure is the target of inductively oriented extraction and rotation algorithms. Because all observed variables (items) may freely load on all factors, so then factors should be rotated to maximize the magnitude of primary loadings and minimize the magnitude of cross-loadings.

3. IN SEARCH FOR THE OPTIMAL NUMBER OF RESPONSE CATEGORIES OF LIKERT SCALE

Considering the assumptions which underline the essence of factor analysis, we should focus now on the extent to which, four distinct types of Likert scales (i.e. 5, 7, 9 and 11) may affect the performance of the factor analysis¹. In practice of social research, Likert scale is typically composed of the battery of items (Shaw, Wright 1967). The vast majority of rating scales and related psychometric instruments were used on 5 or 7 response categories. A smaller number of categories, based on 2 or 3 responses should be considered as insufficient (Cox 1980). On the other hand, increasing the span of the scale from 3 to 9, or even up to 11 points, might also bring some negative effects. For instance, too wide range of scale affects the perception of respondents, what limits their chances of giving the correct answer on particular item. According Stobiecka (2003), a man can use a maximum 11-point scale. The human ability to differentiate between various response options in this regard is very limited. These restrictions appear even faster at scale where not all response categories are described in words (i.e. labeled). As Sztabiński explained (2003: 165) “the researcher's need to use too long scale (e.g., 9 point) causes in the mind of the respondent, first of all, a tendency to flatten scale, which means combining various points on the scale. The attempt to interpret and respond to scale comes in the next phase, far too late.”

In literature, the search for the optimal number of response categories on Likert scale was also considered in context of the reliability level in the measurement to be achieved on the basis of collected responses, as derivatives of the used measurement scales (see e.g. the work of Tarka, Kaczmarek 2013;

¹ More information about the procedure and conditions of constructing Likert scale can be found in his work (Likert 1932).

Alwin 1997; Grigg 1980; Cicchetti, Showalter, Tyrer 1985; Givon, Shapira 1984; Schutz, Rucker 1975; Matell, Jacoby 1971). In most of these works, an increase in the level of reliability (due to transition from a 5-, to 7-point scale) was confirmed. Simultaneously, as Churchill and Peter (1984) admitted, above a certain limit (then unknown) this increase does not compensate for the randomness of responses, which may have a negative impact on the reliability of measurement results.

In yet other comparative studies, validity was also used as a criterion for judging the performance of scales with different numbers of response categories. For instance, Matell and Jacoby (1971) carried out a thorough empirical study comparing scales with varying numbers of response categories (from 2 to 19) and concluded that even 2 response categories may be adequate in practice. They suggested that validity is independent of the number of response categories, and their results implied that collapsing data from longer scales into 2-point or 3-point scales would not diminish the validity of the resulting scores. However, Loken et al. (1987) examined the criterion validity of various scales through their ability to differentiate between different population groups and found 11-point scales to be superior to 3-point or 4-point ones. Also Hancock and Klockars (1991) found that 9-point scale scores correlated better than 5-point scale scores with objective measures of the original stimuli.

4. THE EMPIRICAL STUDY AND RESULTS

In determining the influence of different types of response categories of the Likert scale items used in EFA, the author conducted a comparative analysis. Data was collected in 2014 on the basis of a random sample of 200 first – year students (aged between 19–21) in selected universities in Poznan such as: Adam Mickiewicz University, University of Technology, University of Economics, University of Life Sciences and University of Medical Sciences. The subject of the study was the attitudes of young people towards unethical activities of companies in the market. During the study, participants were asked to express their attitudes to items measured on 5-, 7-, 9- and 11-point Likert scale.

For each scale (5-, 7-, 9- and 11), the same statements within particular items were used. Subsequently, each of the four variants of the scales (see below), along with the equivalent items was introduced to four separate questionnaires. After that, each questionnaire was delivered to the respondents assuming two-week interval time. In labeling the response categories, the following criterion of answer codes (stressing mainly marginal ends of scale) was used: **5 scale:** [1] – *totally disagree*, [3] – *neither disagree, nor agree*, [5] – *totally agree*; **7 scale:** [1]– *totally disagree*, [4] – *neither disagree, nor agree*, [7] – *totally agree*; **9 scale:** [1]– *totally disagree*, [5] – *neither disagree, nor agree*, [9] – *totally agree*; **11 scale:**

[1]– *totally disagree*, [6] – *neither disagree, nor agree*, [11] – *totally agree*. After data collection, by using a principal axis factors extraction method, two latent factors were discovered. They were given the following names: **poor quality products (PQP)** and **lack of social responsibility (LOSR)**².

In running this study the author mainly took an interest in the extent to which, 5-, 7-, 9- and 11-point scale would affect the extraction of factors. Hence, the following stages of extraction were assumed. They pertained to calculation of the: item communalities; total variance explained, matrices of rotated factors along with their coefficient reliabilities; Kaiser-Meyer-Olkin measure of sampling adequacy test and Bartlett's test of sphericity.

Now, having based on two-factor model and comparing the results derived from Table 1, we can say that most of the communalities regardless of the type of used scale, obtained sufficient levels³. However, they seem to be much higher in 9 and 11 point scale as compared to alternatives, i.e. 5 and 7. Although these communalities are not as high as it might be expected (e.g. 0,90) they are not weak, either. If they were very low, then we should engage with another factor and so on. The weakest communalities are represented by 0,50 cut-off level in 7 scale and the following items: X_5 (0,324), X_6 (0,417), X_{10} (0,451).

Table 1. Item communalities after extraction in reference to 5-, 7-, 9-, 11-point scale based on two factors

Items	After extraction			
	5-scale	7-scale	9-scale	11-scale
X_5	0.511	0.324	0.536	0.720
X_6	0.545	0.417	0.689	0.674
X_7	0.593	0.521	0.620	0.709
X_9	0.647	0.531	0.636	0.716
X_{10}	0.533	0.451	0.713	0.726
X_{11}	0.685	0.513	0.737	0.795

Source: own construction

Legend: Shaded color denotes poor communality of i -th item.

² Factor (PQP) was loaded with observed variables (items) as: X_5 – *In recent years, the quality of products offered by the companies has not been improved*; X_6 – *Companies do not take an effort to design products that meet their real needs*; X_7 – *Companies produce products that rapidly wear out, thus cashing in on client servicing*. Factor (LOSR) loaded with items such as: X_9 – *Companies do not pay much attention to the assumption that the client represents the most important element in their activity and business*; X_{10} – *Companies are more interested in pursuing their profits than the overall good of clients*; X_{11} – *Companies in chasing the clients, have changed their marketing practices and impact on clients for worse*. These two factors were part of multidimensional theoretical construct that has been given the name *Unethical Marketing Operations of Companies on Market*.

³ Communalities indicate the amount of variance in each variable that is accounted for. Small values indicate that items do not fit well to the model and should be dropped from further analysis.

In the next phase of factors extraction one referred to total variance explained (Table 2). Also here one can notice differences between 5, 9, 7 and 11 scale. In two cases (i.e., except 5 and 7 point scale) two-factor model was nearly explained at 60% variability within the set of six items. Hence, one can reduce the complexity of the data after rotation by using only two factors on the following levels of information loss (41,923% in 5 scale; 46,604% in 7 scale; 38,637% in 9 scale; 23,554% in 11 scale)⁴. If there are more response categories added on the scale, a greater percentage of the total variance is explained. However, the weakest level of information in the context of the variance explained is provided by 7 scale.

Table 2. Total variance explained in reference to 5-, 7-, 9-, 11-point scale based on two factors

Factors	Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
5-point scale						
1	2.173	36.225	36.225	1.926	32.092	32.092
2	1.311	21.852	58.077	1.559	25.985	58.077
7-point scale						
1	1.986	33.097	33.097	1.613	26.889	26.889
2	1.218	20.299	53.396	1.590	26.507	53.396
9-point scale						
1	2.356	37.273	37.273	2.008	33.722	33.722
2	1.887	24.090	61.363	2.103	27.641	61.363
11-point scale						
1	3.425	48.423	48.423	2.976	40.051	40.051
2	2.130	28.023	76.446	2.510	36.395	76.446

Source: own construction.

The rotated matrices of factors determine the representativeness of particular factors, assuming four levels of measurement (5, 7, 9, and 11). From Table 3 we notice that, values of the respective factor loadings appear to be different, too. However, the first and second factor in 7-point scale is largely unaffected by the rotation and loadings are more difficult to interpret, especially in items X_5 , X_6 and X_{10} . In fact, they indicate cross-loadings, what can be partially confirmed by somewhat unclear size of the Alpha coefficient reliability⁵ calculated for every factor F_1 and F_2 . It was given as follows:

⁴ These differences were due to subtractions of cumulative percentage of the respective factors from 100%.

⁵ Cronbach (1951) created the formula for reliability test, that was derived (as a source) from K-R20 Kuder and Richardson's earlier formula.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum_{i=1}^k \sigma_{X_i}^2}{\sigma_X^2} \right) \quad (3)$$

where: k – number of items in the scale, where $k \geq 2$, $\sigma_{X_i}^2$ – variance of i -th item, σ_X^2 – total variance of the scale.

More importantly, in case of all examined scales, Alpha shows an increase in values as much as there appears greater spread of response categories. Values for the first extracted factor F_1 are ascending, starting from 5 scale ($\alpha_{F_1, 5 \text{ point}} = 0.557$), next 9 scale ($\alpha_{F_1, 9 \text{ point}} = 0.691$), and finishing with 11 scale ($\alpha_{F_1, 11 \text{ point}} = 0.783$). The same order, (with much weaker reliability coefficients) appears in factor F_2 where Alpha explains internal consistency of the measured items, respectively: ($\alpha_{F_2, 5 \text{ point}} = 0.499$); ($\alpha_{F_2, 9 \text{ point}} = 0.659$); and ($\alpha_{F_2, 11 \text{ point}} = 0.730$)⁶.

Table 3. Matrices of rotated factors and their reliabilities in reference to 5-, 7-, 9-, 11-point scale

Extracted factors and scale types								
Items	5-point scale		7-point scale		9-point scale		11-point scale	
	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
X_5	-0.270	0.547	-0.226	-0.329	0.010	0.649	-0.259	-0.722
X_6	0.124	0.581	0.351	0.525	0.011	0.654	0.104	0.670
X_7	0.228	0.554	-0.171	0.526	0.033	0.630	-0.082	0.709
X_9	0.663	0.104	0.519	0.245	0.696	0.181	0.723	0.123
X_{10}	0.569	0.292	0.461	-0.324	0.726	-0.105	0.696	0.209
X_{11}	0.710	-0.033	0.100	0.461	0.729	0.024	0.792	-0.073
Alpha α	0.557	0.499	0.401	0.394	0.691	0.659	0.783	0.730

Legend: Shaded color denotes relation of i -th item and j -th respective factor. Some of them represent cross-loadings.

Source: own construction.

The last table presents scores that indicate the suitability of data within the structure item in EFA. The Kaiser-Meyer-Olkin measure of sampling adequacy is a statistic that indicates the proportion of variance in item that might be caused by underlying factors. High values (close to 1,0) generally indicate that a factor

⁶ We need to stress at this point, that if Alphas are lower than 0,50 they are rather unacceptable.

analysis is useful with the data at hand⁷. The KMO formula is given as follows (Kaiser 1974):

$$KMO = \frac{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2}{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2 + \sum_{i \neq j} \sum_{j \neq i} \hat{r}_{ij}^2} \quad (4)$$

where: r_{ij}^2 – the element of correlation matrix, i.e. correlation between i -th and j -th observed variable, \hat{r}_{ij}^2 – partial correlation coefficient between i -th and j -th observed variable.

On the other hand, Bartlett's test of sphericity verifies the hypothesis that correlation matrix is an identity matrix, which would indicate that items are unrelated and therefore unsuitable for structure detection. This test is based on the following statistic (Bartlett 1954). Small values (less than 0.05) of significance level in test indicate that a factor analysis may be useful with data.

Now, considering two-factor model, both KMO measure and Bartlett's test confirm the adequacy of data and factors extraction on the basis of four different scales. However 5- and 7-point scale yields the weakest scores. The other two scales, namely, 9 and 11 scale exceeded the threshold of poor structure of data recommended by Kaiser (1974). In conclusion, the wider range of scale, the better final score in KMO and Bartlett's test.

Table 4. Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity based on two factors

	5-scale	7-scale	9-scale	11-scale
KMO	0.597	0.503	0.626	0.763
Significance of Bartlett's test of sphericity	0.000	0.000	0.000	0.000

Source: own construction.

5. CONCLUSIONS

The rating scales that yielded the least reliable scores and brought little sense in the extraction of factors turned out to be those with the response categories such as 5 and 7. The most promising results were derived from scales measured with 9 and 11 responses. These findings provide a suggestion that reliability and simultaneously factors extraction process is not independent of the number of

⁷ Kaiser (1974) accepted the following *KMO* levels [0, 1] where: 0.9 – excellent (meaning a very good structure correlation matrix as an input data); 0.8 – recommendable; 0.7 – decent; 0.6 – average and 0.5 – indicating poor structure.

response categories. Probably, nothing will be gained when researchers use scale with 5 answers. As there are more responses added to the scale, a reliability of the factors, as well as the performance of factors extraction tend to increase. To sum up, the findings suggest that scales with 9 or 11 categories are better as they allow the respondents to express their feelings, attitudes or opinion more adequately.

REFERENCES

- Alwin D.F. (1997), *Feeling thermometers vs. 7-point scales*, "Sociological Methods and Research", vol. 25, no 3, pp. 318–351.
- Bartlett M.S. (1954), *A note on the multiplying factors for various chi-square approximations*, "Journal of the Royal Statistical Society", 16, Series B, pp. 296–298.
- Churchill G., Peter P. (1984), *Research design effects on the reliability of rating scales – a meta-analysis*, "Journal of Marketing Research", no. 4, pp. 360–375.
- Cicchetti D.V., Showalter D., Tyrer P.J. (1985), *The effect of number of rating scale categories on levels of inter-rater reliability: a Monte-Carlo investigation*, "Applied Psychological Measurement", vol. 9, pp. 31–36.
- Cox E.P. (1980), *The optimal number of response alternatives for a scale: a review*, "Journal of Marketing Research", vol. 17, pp. 407–422.
- Cronbach L.J. (1951), *Coefficient alpha and the internal structure of tests*, "Psychometrika", vol. 16, pp. 297–334.
- Givon M.M., Shapira Z. (1984), *Response to rating scales: a theoretical model and its application to the number of categories problem*, "Journal of Marketing Research", vol. 21, November, pp. 410–419.
- Gorsuch R.L. (1974), *Factor analysis*, Saunders Company, London.
- Grigg A.O. (1980), *Some problems concerning the use of rating scales for visual assessment*, "Journal of the Market Research Society", vol. 22, no. 1, pp. 29–43.
- Hancock G.R., Klockars A. J. (1991). *The effect of scale manipulations on validity: targeting frequency rating scales for anticipated performance level*, "Applied Ergonomics", vol. 22, pp. 147–154.
- Kaiser H.F. (1974), *An index of factorial simplicity*, "Psychometrika", vol. 39, pp. 31–36.
- Kendall M.G. (1950), *Factor analysis*, *Royal Statistical Society*, vol. 12, pp. 60–73.
- Kim J.O., Mueller C.W. (1978), *Introduction to factor analysis – what it is and how to do it*, Sage University Press, Newbury Park.
- Likert R., (1932), *A technique for the measurement of attitudes*, "Archives of Psychology", vol. 140, pp. 5–55.
- Loken B., Piri P., Virnig K. A., Hinkle R. L., Salmon C.T. (1987). *The use of 0-10 scales in telephone surveys*, "Journal of the Market Research Society", vol. 29, no 3, pp. 353–362.
- Matell M.S., Jacoby J. (1971), *Is there an optimal number of alternatives for Likert scale items? – study: reliability and validity*, "Educational and Psychological Measurement", vol. 31, pp. 657–674.
- Mulaik S.A. (1990), *Blurring the distinctions between component analysis and common factor analysis*, "Multivariate Behavioral Research", vol. 25, no. 1, pp. 53–59.
- Mulaik S.A. (1987), *A brief history of the philosophical foundations of exploratory factor analysis*, "Multivariate Behavioral Research", vol. 22, pp. 267–305.
- Schutz H.G., Rucker M.H., 1975, *A comparison of variable configurations across scale lengths: an empirical study*, "Educational and Psychological Measurement", vol. 35, s. 319–324.

- Shaw M.E., Wright J.M. (1967), *Scales for the measurement of attitudes*, McGraw-Hill, New York.
- Stobiecka J. (2003), *Porządkowy czy interwałowy charakter skal szacunkowych w naukach społecznych* [w] Gatnar E. (red.), *Analiza i prognozowanie zjawisk rynkowych o charakterze metrycznym*, Wydawnictwo AE w Katowicach, s. 84.
- Sztabiński F. (2003), *Logika badacza i logika respondenta – problem adekwatności narzędzia badawczego*, „ASK, Społeczeństwo, Badania i Metody”, nr 12, s. 165.
- Tarka P., Kaczmarek M. (2013), *Analiza porównawcza metod pomiaru postaw respondentów*, „Wiadomości Statystyczne”, nr 8, s. 37–47.

Piotr Tarka

SKALA LIKERTA A ZAKRES KATEGORII ODPOWIEDZI W PROCESIE EKSTRAKCJI CZYNNIKÓW W MODELU EFA

Streszczenie. Celem artykułu jest analiza porównawcza skali Likerta o różnej podstawie rozpiętości kategorii odpowiedzi: (5, 7, 9 i 11 stopniowej), w kontekście procesu wyodrębniania czynników w eksploracyjnej analizie czynnikowej (EFA). Poruszany w artykule problem dotyczy przede wszystkim wątków metodologicznych zarówno w kwestii wyboru optymalnej liczby kategorii w sferze mierzonych pozycji składających się na skalę Likerta jak i rozpoznania zmian, różnic lub podobieństw towarzyszących (w wyniku oddziaływania czterech typów skal) procesowi wyodrębniania i określania odpowiedniej liczby czynników w modelu EFA.

Słowa kluczowe: eksploracyjna analiza czynnikowa, skale 5, 7, 9 i 11 punktowe, badania eksperymentalne, marketing.