

Measurement method by finding the global weight of item

"A theoretical and practical model"

Ph.D. Assistant Professor Sattar Jabbar Ghanem

Department of Psychology / College of Arts / University of Mustansiriyah

ABSTRACT

The researchers tried to reach the method that has the highest accuracy in psychological measurement by eliminating the defects in previous and common measurement methods such as Likert's method or the Thurstone method, so the current research aims to analyze both methods in an attempt to arrive at the method that is believed to be the most appropriate for measurement. Therefore, we find many psychologists build their tools to measure the psychological concept that they want to study and understand, and in this case verification of the validity of the scale construction, therefore the degrees obtained from it accurately represent the quantity of the concept or the psychological characteristic of the individuals of the sample and it is not an easy task, on this basis the research offers two hypotheses: 1- A weight can be given to item in the scale based on its weight among the members of the community, 2-The statistical method represented by the factor analysis can be used to find the weight of item, In a series of innovative procedures in the first application stage of the scale and the application of the factor analysis, the score obtained by the subject on the scale was calculated using the procedures followed in Likert's method and re-correcting each form by giving a special weight to each item in order to give each item a weight consistent with its ability to represent the psychological concept, It is a matter closer to accuracy and objectivity in measurement as this weight is obtained by the researcher through the process of factor analysis, and thus the weight of item is the value of its saturation with the psychological concept, This weight is divided by the number of answer alternatives for each item, and the outcome of this process is multiplied by the previous weight of the alternative, this is what provided support and attribution to the research hypotheses.

KEYWORDS: psychometrics, factor analysis, construct validation, item Weight

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

The problem of research:

After researching the methods of psychological measurement and its stages of development in the psychological and educational sciences, starting with the method used by Bogardus, the researcher believes that researchers who preceded him all tried to reach the method that guarantees the highest accuracy in measurement by eliminating the defects in the previous measurement methods As will be mentioned later. therefore, the researcher believes that both Likert and Thurstone method are good methods of measurement, but each of them has its flaws and advantages. Accordingly, the researcher seeks to analyze both methods in an attempt to arrive at the method that he believes is the most appropriate for measurement innovated by the researcher.

Introduction

After four years of continuous work, two halves of a large bridge across the Rhine have met to connect Germany and Switzerland. The engineers working on the bridge were astonished to find that there was a difference of 21 inches in height between the two sides: Various measurements of sea level (North Sea versus Mediterranean) were

used. Clerk Maxwell says: " We owe all the great advances in knowledge to those who endeavour to find out how much there is of anything. And Lord Kelvin says: "One's knowledge of science begins when he can measure what he is speaking about, and express it in numbers. Johannes Clare says "As the ear is made to perceive sound and the eye to perceive colour, so the mind of man has been found to understand not all sorts of things, but quantities. It perceives any given thing more clearly in proportion as that thing is close to bare quantities as to its origin, but the further a thing recedes from quantities, the more darkness and error inheres in it."Johannes Kepler, Opera, (1595) (Eysenck, 1973).

Measurement problems can occur (sometimes with huge consequences) as in the bridge problem, as in psychology we do not disagree but rather our task is much more difficult. Because what we want to measure is an abstract concept and not a physical thing that can be measured with the physical tools of measurement such as height and weight. From here we find that in measurement we convert concepts into quantities, this process is a risky daring. Edward L. Thorndike (1874-1949) wrote in his book, Educational Psychology: "We overcome

facts of nature when we observe and experience them. And when we measure them, we make them our servants" (1903, p. 164). Thorndike's focus on measurement and quantification was just as important as experimental psychology, poorly prepared by schools in mathematics and largely self-education in statistics, Thorndike was a guide to psychological and educational measurement and use of the universal language of science which is the language of numbers. o he said: "Whatever exists, exists in an amount (amount) and what exists in an amount (by amount) can be measured " this saying includes knowing the quantity, quality and accuracy in knowing the amount of this quantity (Eysenck, 1973).

Therefore, we find many psychologists build their tools to measure the psychological concept that they want to study and understand. In this case, verifying the validity of the scale construction, and the grades obtained from it accurately represent the quantity of the concept or the psychological characteristic of the sample members, which is not an easy task. This difficulty is not only accurate in measuring psychological structures, but it is also complex. It is relatively easy to determine which sea should be a criterion for calculating height above sea level, but accurately and correctly quantifying a psychological variable is a great challenge.

Although there are different ways to convert the psychological variable into quantitative numbers, but it remains imprecise when reviewing its procedures because it includes different behaviors and perceptions expressed in items that together constitute the psychological scale, therefore different behaviors that express the psychological concept cannot take the same weight in measurement as Likert did, and as Thurston emphasized.

This research will shed light on item weight in psychometrics, It explains why the equal weight of the scale items establishes a serious and unacknowledged threat to the efforts of researchers in psychological and educational sciences, and concludes with a new measurement method presented by the researcher to address the problem of item weight.

Research hypotheses :

- 1- A weight can be given to item in the scale based on its weight among the members of the community.
- 2-The statistical method represented by the factor analysis can be used to find the weight of item.

Research literature

• Thurstone Method (1929)

Since 1929, Thurston presented his theoretical analysis in an attempt to develop a method for measuring attitudes based on the basic ideas of psychophysics. It depends on creating a balance between the severity of the psychological phenomenon (attitude) and the physical units of measurement (the degree on the test). This method aims to choose items to measure attitudes that represent different degrees of attitude in such a way that accurate values can be determined for each of them .Thurston has used this method in measures of attitudes towards a number of subjects. In this case, the scale consists of a number of statements, including equal intervals or distances, describing the attitude from maximum positivity to maximum negativity, this Thurston method is called "Equal Appearing Intervals" (Guffey,etal. ,2007; Stevens, 1946).

On the method of preparing scale and estimating weight of each statement, test designer asks a large number of individuals to present some sentences describing their attitudes towards the subject to be measured, Subsequently, these sentences are reviewed and crystallized through literature on the subject to be measured and reformulated in the form of a clear and short sentence .Thus, a large number of statements and phrases are collected , possibly more than a hundred ,Thurston believes that it measures the attitude he wants to measure, it represents the extent of approval or acceptance to rejection or aversion ,Then each statement is written on a separate card and the statements are presented to a group of expert referees in the field (their number may reach a hundred),Experts are asked independently to place each statement in a category ranging from 1 to 11, so that the most positive expressions are in Category No. (1) and the most negative are in Category No. (11). It is noted that role of experts in this case is not to express their situation or attitudes towards the subject of measurement or to choose statements that express their attitudes , but rather that their task is only to determine the extent to which these statements express the attitude to be measured without regard to their personal situation (Likert, & Murphy ,1993; Guffey,etal. ,2007).

In the next step, the weight of each statement is obtained by calculating the median of the experts' answers for each statement ,Then the percentiles are determined and the statement are chosen so that their percentiles are graduated in a manner corresponding to gradient from 1 to 11, the statement that occupies the first place is in the first percentile, it is the phrase that expresses the most positive attitudes ,the statement in 11th position

that expresses the most negative trends. After this process, the ambiguous or inappropriate statements about which the experts disagreed are excluded. Whereas, the expressions that they have unanimously accepted are accepted according to the test of homogeneity that appeared in the answers of the experts by calculating the Semi-Quartile Range* for each statement. Considering that Semi-Quartile Range is one of the measures of dispersion that is best used in cases where we use the median as a measure of central tendency. statements that get half of the low quadrant range are accepted as an indication of the severity of the homogeneity. After determining statements that are valid for measurement, the respondent's score on the scale is the value of the mean calculated from the sum of the statements he answered. (Likert, & Murphy, 1993; Guffey, et al., 2007; Edwards & Kenney, 1946)

• Likert Method (1932)

Likert, R., devised a method for measuring psychological attitudes in an effort to overcome the difficulty of experts in the field in the Thurston method. Today, this method is considered the most common method in use to measure the psychological attitudes towards a certain phenomenon due to its ease of application, correction and extracting its results (Stevens, 1946; Edwards & Kenney, 1946)

Thus Likert dispensed with the expert judgment that was one of the conditions of Thurston, as scale consists of a set of positive and negative items that are equal in number. He has collected a large number of statements from a wide range ranging from extremist or moderate approval to moderate or extremist opposition with the exclusion of statements that he thinks are neutral. He tested these statements on a group of individuals asking them to respond to each statement in one of the following categories (I strongly agree, I tend to agree, I am unsure, I tend to oppose, I strongly oppose). And then correct responses to these statements that indicate support by giving the previous categories the following values (5,4,3,2,1) respectively, while the opposing response gives the values contrary to the previous order (Edwards & Kenney, 1946)

The Likert method is described as a model that achieves effectiveness, adequacy and accuracy in measurement (Borg & Call, 1975). It has a number of advantages, such as:

- a. Ease of construction and correction .
- B. Achieve a more consistent scale.

C. It allows the respondent to indicate the degree or intensity of his acceptance or rejection of the single item.

D. does not require a large number of experts when used.

e. Reliability tends to be good, due in part to the large range of responses allowed by the respondent. (Oppenheim, 1973; Stevens, 1946; Edwards & Kenney, 1946).

From this review, we find that the Thurston's method is distinguished by the fact that it gives each item its own weight on attitude scale. This is closer to accuracy than in Likert's method, which gives all items the same weight. However, obtaining an adequate number of experts constitutes a hindrance in using this method of measurement, as well as "the use of biased or extremist experts affects the design and construction of the scale, we also find that the respondent on the Thurston style only answers item that applies to him, while Likert's method is characterized by placing a number of alternatives in front of each item, to which the weight of the item is distributed equally, the respondent is asked to express his attitude towards each item of the scale, and for this reason, each item provides us with information about the respondent. therefore, the researcher will employ the logic on which each method is based. He uses the procedures in Likert's method of measurement, then completes them by giving a special weight for each item, this weight is obtained through the process of factor analysis by making weight of item is value of its loading with the psychological concept after process of rotating of axes. this weight is divided equally among alternatives of item, and as shown in the steps for building social distance scale by researcher.

Scale application and sample

social distance measure consisting of (23) items was chosen, which was previously prepared by the researcher in a previous study, and researcher applied it to a random sample of university students of (413) male and female .

factor analysis

Factor analysis is used to find out whether the elements on the scale have similar patterns of responses. Are these elements "combined" to create a structure? So the basic assumption of factor analysis is that for a set of observed

variables there is a set of basic variables called factors (the least number of elements or scale items), which can explain the interrelationships between those items. So the question is: do all these items measure what we call social distance? Factor analysis also answers the following questions: if scale is composed of domains (factors), are these domains valid? if there are no fields (factors) within the scale, then can we find fields for this scale, or does it represent one basic field? Are these fields compatible with study population or can other fields be derived that are appropriate to the characteristics of the study population? It is worth noting that factor analysis is a method for modeling observed variables, and their covariance structure, in terms of fewer unobservable latent "factors". Factors are usually seen as general concepts or ideas that may describe an observed phenomenon. Variation is one of the most basic concepts in factor analysis, that is, the amount of dispersion of numerical values from the sample mean. We are looking to understand how the various fundamental factors affect the contrast between our elements (and in this research, each of the scale item is considered an element). Each factor will have an impact, but

some will explain more variance than others, which means that the factor represents more accurately the elements (items) that make it up. (2) So we focus mainly on variance and covariance rather than the middle. Factor analysis assumes that variance can be divided into two types of variance which are common variance and unique variance (Stevens, 1986; Kim, 1978a; Cattell, 1966).

Factual analysis of the social distance scale paragraphs

The factor summarizes the associations between the various items and these correlations constitute structure or formative qualities of the factor and which give him his psychological meaning. The researcher used the Principal Component Method developed by Hotelling (Cattell, 1966; Ferguson & Takane, 1989; Stevens, 1986) on sample of (413) male and female from university students, scale items (23) represented a variable that was used in factor analysis process. Table (1) shows the basic statistics for these variables.

Table (1)

Basic statistics for the paragraphs of the social distance scale

items	middle	standard deviation	items	middle	standard deviation
1	5.69	1.99	13	4.04	2.10
2	4.82	1.84	14	3.25	2.29
3	4.65	1.92	15	3.84	2.19
4	3.86	2.22	16	2.02	1.66
5	5.68	2.01	17	3.16	2.08
6	4.03	2.11	18	4.21	2.23
7	4.74	2.29	19	2.52	1.94
8	4.79	2.25	20	2.54	1.99
9	3.11	2.10	21	2.59	2.16
10	5.79	1.93	22	1.92	1.79
11	4.46	2.25	23	1.68	1.59
12	2.06	1.61			

Pearson correlation coefficients were calculated between these variables and Correlation Matrix was formed on which the factor analysis was performed. The analysis process produced (23) factors arranged in descending order according to their contribution to calculated subscriptions. The commonly used rule in the basic components method is that the contribution of factor to sum of the social contributions is not less than one

(Ferguson & Takane, 1989; Kim, 1978a). Accordingly, Table (2) indicates five factors that were obtained according to this rule. We also note that Table (2) indicates of loading each factor with (23) variable (item) which means degree of item is linked to factor, the rows in base of the table show (distinct root) eigen value of resulting factors, i.e. contribution of each factor to sum of the contributions, and variance ratio of each factor to

total variance of the correlation coefficients matrix, which is called explained variance or percentage of the contribution, the last row in the table shows combined variance.

Table (2) shows that correlational variance of five factors resulting from correlation matrix analysis process represents about (55.681) of total variance, this is the maximum correlational variance extracted from the correlation matrix, and data in Table (2) are not flexible for any

psychological interpretation, therefore, the researcher deliberately rotated assignment axes to new positions, as rounded factors would be a description of structural vector characteristics (Ferguson & Takane, 1989), this means obtaining the largest number of high positive loadings and loadings close to zero, and thus we obtain new derived factors and the meaning is assigned to them, i.e. redistributing loading values so that sub-groupings of the factors are clear.

Table (2) : Factors resulting from Social Distance Scale items Correlation Matrix Analysis (Before Rotation)

Items	Factors				
	1	2	3	4	5
1	0.486	-0.301	0.216	0.283	9.892E-02
2	0.257	5.549E-03	-0.424	0.699	9.152E-02
3	0.303	4.874E-02	-0.367	0.732	0.140
4	0.673	-6.071E-02	7.717E-02	0.104	-0.271
5	0.591	-0.274	0.357	8.277E-02	-5.881E-02
6	0.636	1.839E-02	9.984E-02	3.633E-02	-0.210
7	0.686	-0.181	0.158	-6.085E-04	-0.297
8	0.670	-0.235	0.158	0.103	-0.226
9	0.565	0.118	-3.278E-02	-9.232E-02	0.160
10	0.518	-0.232	0.434	1.282E-02	0.206
11	0.637	-0.222	0.243	-2.019E-02	0.187
12	0.576	0.249	-0.340	-0.178	0.265
13	0.480	0.148	-0.186	-9.657E-02	-0.457
14	0.586	0.193	-0.349	-0.215	-0.143
15	0.528	0.216	-0.356	-0.101	-0.224
16	0.500	0.341	-0.332	-0.277	0.201
17	0.615	0.195	-8.636E-02	-0.167	0.386
18	0.628	2.896E-02	0.203	-0.148	0.376
19	0.167	0.499	4.586E-03	-6.048E-02	-0.163
20	-5.890E-02	0.559	0.271	0.281	-1.476E-02
21	-2.233E-02	0.669	0.298	0.166	9.908E-02
22	-1.148E-02	0.700	0.312	8.710E-02	-7.211E-02
23	2.789E-02	0.727	0.301	8.431E-02	-6.340E-02
Contributio n of factor	5.751	2.726	1.710	1.491	1.128
Explained	25.003	11.854	7.435	6.482	4.906

variance					
Clustered variance	25.003	36.858	44.293	50.775	55.681

it is worth noting that the researcher used in process of rotation the Varimax method to maximize normal variance of rotation, which was proposed by Kaiser and rotation of factors to make Varimax criterion of normal varimax the largest possible.

Table (3) indicates factors resulting from recycling process, it is noticed that values of loaded with factors have changed if we compare them with values before rotating of axes. It is noted from the table that the variance ratio of each factor

to the total variance remained constant before and after rotation process. Based on Therston criterion, which emphasizes importance of item being loaded in a way that has practical significance in one factor and weak in others (Child, 1979). To judge the value of the factor loadings of the variables (items) of practical significance for all of this, the loadings is close to zero, if it is less than (0.4) (Stevens, 1986; Kim, 1978b), this is used as a criterion on the basis of which the variables (paragraphs) are accepted in the factor.

Table (3)

Factors resulting from rotating of axes (social distancing measure)

Items	Factors				
	1	2	3	4	5
1	0.626	3.642E-02	-0.106	-1.996E-02	0.241
2	4.669E-02	7.392E-02	-2.985E-02	9.290E-02	0.852
3	0.106	0.107	3.827E-02	5.825E-02	0.870
4	0.544	0.111	1.927E-02	0.470	0.130
5	0.734	3.417E-02	-5.752E-02	0.137	2.432E-04
6	0.495	0.170	7.994E-02	0.419	6.466E-02
7	0.624	7.670E-02	-7.117E-02	0.464	5.218E-04
8	0.653	5.330E-02	-9.871E-02	0.377	9.699E-02
9	0.322	0.478	6.205E-02	0.172	5.330E-02
10	0.712	0.164	-1.494E-02	-0.117	-6.959E-02
11	0.671	0.300	-9.669E-02	2.331E-02	4.311E-03
12	8.819E-02	0.728	1.572E-02	0.232	0.135
13	0.137	0.136	4.733E-02	0.682	1.811E-02
14	8.820E-02	0.495	-2.571E-02	0.560	5.073E-02
15	5.174E-02	0.382	2.006E-02	0.588	0.132
16	-1.156E-02	0.715	8.006E-02	0.274	2.419E-02
17	0.293	0.709	7.773E-02	5.608E-02	5.301E-02
18	0.537	0.547	6.629E-02	-5.849E-02	-6.082E-02
19	-7.246E-02	0.170	0.430	0.294	-4.352E-02
20	-2.612E-02	-7.145E-02	0.668	-5.417E-02	0.116
21	-3.849E-02	8.343E-02	0.745	-0.104	2.822E-02
22	-5.345E-02	2.373E-02	0.768	4.873E-02	-6.804E-02
23	-4.222E-02	6.307E-02	0.785	6.818E-02	-5.739E-02

Contribution of factor	3.786	2.712	2.464	2.196	1.648
Explained variance	16.459	11.793	10.713	9.550	7.167
Clustered variance	16.459	28.252	38.965	48.514	55.681

The commonly used rule in basic components method is that the contribution of the factor to the total contributions should not be less than (1), and this rule is an arbitrary rule (Ferguson & Takane, 1989). The number of factors resulting from factor analysis model depends on estimates of contributions. Therefore, choosing appropriate solution depends on external criteria that are not related to the model. The significance of factor is considered one of subjective criteria for determining one of factors (Kim, 1978b). In small-sized samples, one of factors can be judged to be insignificant, even though it contributes significantly, on the other hand, the least contributing factors in the case of very large samples may be judged to be significant (Stevens, 1986). Among other methods in this field: the extent of conformity with results of other studies in the same field, and weight of loadings of factor, and this may be a criterion of judgmental methods in this field if scientific experience is available in the research subject (Kim, 1978b) so it has been shown that the factors resulting from recycling process (Table 3) have psychological meanings according to significant loadings of items except for the fifth factor, which the researcher omitted due to its ineffectiveness in measuring. As this factor carries the lowest effectiveness to the total contribution of factor, as it reached (1.648), and it includes only two items, which are clear from their observations that they are not scientifically useful "in measuring social distance.

Factor weight method of measurement invented by the researcher

In the first application stage, on basis of which factor analysis was calculated, the score obtained by subject was calculated on social distance scale using procedures followed in Likert method, but at this stage we will re-correct each questionnaire by giving special weight to each item, this procedure is based on the logic: In fact, the items differ in their ability to represent the concept for which they are intended. Since Likert's method gives all items same weight, so giving each item a weight consistent with its ability to represent the psychological concept is a matter closer to accuracy and objectivity in measurement, this weight is obtained by the researcher through process of factor analysis, by making weight of item a consequence of value of its loading with the

psychological concept, this weight is divided by the number of answer alternatives for each item, and the result of this process is multiplied by the previous weight of alternative. Whereas the researcher had placed a seven-scale to evaluate responses on items of social distancing scale, which are: (I desire a lot, I desire a medium degree, I desire a little degree, neutral, I do not desire a little degree, I do not desire a moderate degree, I do not desire completely), through process of factor analysis, the value of loading of the first item with psychological concept was (0.626) (Table 3), and this value represents weight of item in the scale, so this weight is divided by number of alternatives, which is (7), so the result is (0.0894), so this value is multiplied by scale of alternatives degrees (1, 2, 3, 4, 5, 6, 7) respectively, which was previously determined according to Likert's method, so the result will be (0.089, 0.179, 0.268, 0.358, 0.447, 0.537, 0.626), which are degrees that will be used in correction of item according to alternative indicated by participant. thus, each item will be corrected by giving special weight to it, as is case on first item, the sum of weights of alternatives indicated by participant represents value of score obtained on the scale. Since researcher has obtained four factors for social distance scale through process of factor analysis, so each participant will receive four scores representing these factors and each factor consisting of sum of significant items in it. In theory, the lowest score that respondent can get on the four factors starting from the first factor is (0.80, 0.52, 0.49, and 0.45) degrees respectively, and it is the lowest difference in social distance and represents complete acceptance of the other, while the highest score that the respondent can get on four factors starting from the first factor is (5.60, 3.67, 3.40, 3.18), which is the highest difference in social distance and represents the total rejection of the other.

RESULTS

It appears from the above procedures that the research hypotheses:

- 1- Weight can be given to item in scale based on its weight among research community members.

- 2- The statistical method of factor analysis can be used to find weight of item.

These hypotheses were realized and proven.

RECOMMENDATIONS

The researcher recommends using this method in psychological measurement and verifying it on other psychological variables.

REFERENCES

- Borg, W. & Call, M. (1975) . **Educational Research** . 3rd , New York : Longman.
- Cattell, R. B. (1966). **The scree test for the number of factors**. Multivariate Behavioral Research, 1, 245-276. https://doi.org/10.1207/s15327906mbr0102_10
- Child, D. (1979) .**The Essentials of Factor Analysis** . New York : U.S.A.
- Edwards, A. L., & Kenney, K. C. (1946). **A comparison of the Thurstone and Likert techniques of attitude scale construction**. Journal of Applied Psychology,30(1), 72 –83. doi:10.1037/h0062418
- Eysenck, H. J. (1973).**The Measurement of Intelligence** , Medical and Technical Publishing Co. Ltd., Lancaster, England .
- Ferguson, G. & Takane, Y. (1989) . **Statistical Analysis in Psychology and education** . New York : McGraw-Hill .
- Guffey, James & Larson, James & Zimmerman, Loren & Shook, Brenda. (2007). **The Development of a Thurstone Scale for Identifying Desirable Police Officer Traits**. J Police Crim Psychol. 22. 1-9. 10.1007/s11896-007-9001-8.
- Kim, J. O. & Mueller, C. W. (1978a). **Introduction to factor analysis: What it is and how to do it**. Beverly Hills, CA: Sage.
- Kim, J. O. & Mueller, C. W. (1978b). **Factor analysis: Statistical methods and practical issues**. Beverly Hills, CA: Sage.
- Likert, R., Roslow, S., & Murphy, G. (1993). **A simple and reliable method of scoring the Thurstone attitude scales**. Personnel Psychology,46(3), 689 –690. doi:10.1111/j.1744-6570.1993.tb00893.x
- Oppenheim, A. (1973) . **Questionnaire Design and Attitude Measurement** . London : Heinemann .
- Stevens, J. (1986). **Applied multivariate statistics for the social sciences**. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stevens, S. S. (1946). **On the theory of scales of measurement**. Science,103(2684), 677–680.doi:10.1126/science.103.2684.677