Principal Component Analysis applied to survey data: Methodological aspects and application

Bourass Hind

Université Mohammed V Rabat, FSJES Agdal

bourasshind1@gmail.com

Soussi Noufail Outmane

Université Mohammed V Rabat, FSJES Salé

soussioutmane@gmail.com

ORCID : 0000-0002-0269-7935

Abstract—

The objective of this paper is to highlight the methodology of principal component analysis. It is a branch of multivariate descriptive statistics that allows for the simultaneous processing of any number of quantitative data. Firstly, it is interesting to review the different methods used for the construction of the questionnaire and the necessary steps for its elaboration. Secondly, the techniques used can be generalised to two complementary analyses: reliability analysis and dimensionality analysis. Secondly, the focus will be on assessing the internal consistency of the scales. It is therefore appropriate to consider principal component analysis as an extremely powerful tool for synthesising the information contained in the various data to have a representation that allows easier interpretation.

Keywords—

Factor analysis, Survey, Principal component analysis, Multivariate statistics

I. Introduction

In management science, there are several methodological choices in applied research. Surveys are one of the most widely used research methods in this discipline, allowing the analysis of different types of variables, both qualitative and quantitative.

The methodological choice based on quantitative studies makes it possible to obtain numerical

relational information that serves to better understand the chosen theme or subject. Moreover, recent developments in quantitative tools make it possible to quantify qualitative variables, and one can, conversely, make interpretations and draw qualitative conclusions from quantitative data (Barbet, 1988).

The aim of this paper is to highlight a methodology often used by management science researchers whose main research instrument is data drawn from questionnaires. This is the methodology of exploratory analysis of measurement instruments. Exploratory factor analysis aims to explore a number of items in a given population. The techniques used can be generalised to two complementary analyses: reliability analysis and dimensionality analysis. For this purpose, a statistical analysis is used: factor analysis.

This paper is divided into four points. The first point describes the steps involved in designing the questionnaire. The second point describes the choice of a scale to measure a latent variable from a set of items. The methodology for the evaluation of these measurement instruments is discussed with a numerical application. The last point is reserved for the methodology of the application of the principal component analysis method and its statistical evaluation

II. Methodology for the construction of the questionnaire

A. *Definition and construction of the research instrument*

The questionnaire is a document on which the answers or reactions of one or more individuals are noted. Indeed, it is applied to a set (sample) that must allow statistical inferences to be made, in order to measure and evaluate behavioural phenomena in a population. Therefore, the systematic implementation of pre-test procedures and scientific design of questionnaires is essential for the quality of the data collected, in particular to minimise so-called measurement errors.

The questionnaire survey is considered as a data collection tool to understand and explain phenomena on the one hand, and to study the characteristics of a population on the other. In addition, the questionnaire may allow us to verify, confirm or refute hypotheses related to a research.

International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

Thus, an essential step in market research. Firstly, it provides the data needed to create scales for the brief as its structured form makes it easier to analyse the hypotheses. (Malhotra, 2008).

Secondly, it is a measurement tool. Its main purpose is to operationalise the user's request for information in a format that allows statistical measurement. The concepts of 'reality' must be operationalised in such a way that specialists and users can perform the necessary analyses, which the questionnaire designer can implement in the questionnaire, and which the respondents can understand and answer correctly.

However, in order for the questionnaire to fulfil its functions, it was necessary to have a clear and precise definition of the questionnaire. The first phase of the development of the questionnaire was a very crucial phase for us in order to be in line with the scientific approach. For this purpose, we were largely inspired by the steps proposed by (Churchill, 1998):

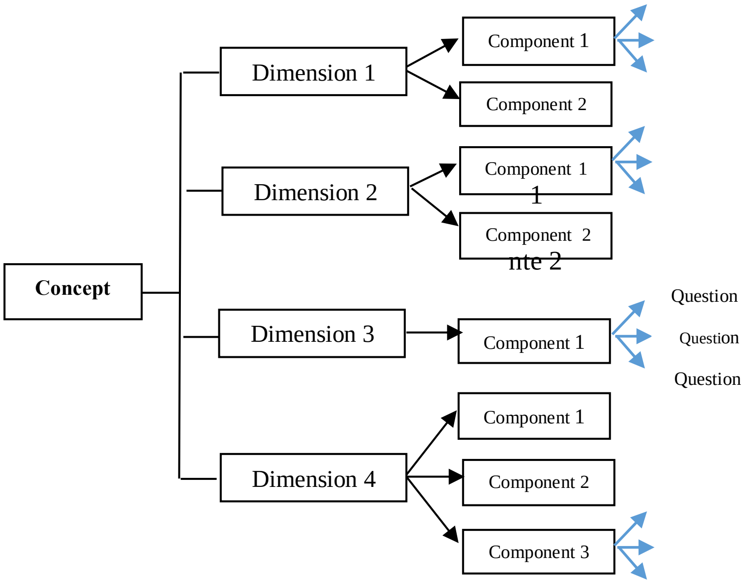
- Defining the purpose of the research ;
- The specification of the information sought, i.e. the objectives and hypotheses to be tested;
- The type of questionnaire and its mode of administration ;
- The content of individual questions ;
- The form of response to each question ;
- The wording of each question ;
- The sequence of questions ;
- The physical characteristics of the questionnaire;
- Pre-testing of the questionnaire, and drafting of the final version.

The latter plays a key role in the data collection process, as well as influencing the image of the statistical agency that uses it. It also. It has a significant influence on respondent behaviour, interviewer performance, collection costs and respondent relations, and therefore has a considerable impact on data quality.

The process of designing the questionnaire consists of several successive steps:

- The development of a conceptual framework,
- The writing and sequencing of questions,
- The correct use of visual design elements and the technical implementation of electronic questionnaires.

Figure 1: Construction of a questionnaire



Therefore, we will now deal with the essential steps related to the construction and implementation of the questionnaire.

B. Forms and phases for the construction of the questionnaire

The questionnaire survey is a reliable measurement instrument that is easy to[1] administer and analyse (Malhotra, 2008).
Questionnaires can be administered by interviewers in person or by telephone, or they can be self-administered on paper or in another medium, such as an audio cassette or the internet. Respondents may be asked to report on themselves, others in their household or other entities, such as businesses.

The design of the questionnaire, in accordance with the Code of Practice [2] . Therefore, the wording, structure and presentation of all questionnaires must lead to valid and reliable results. There are two types of questionnaires:

- Direct-administration questionnaires: in this case, the respondent writes his/her answers on the questionnaire. The person conducting the survey (the interviewer) may or may not be present; if he or she is present, he or she may clarify the content of an answer if the respondent so requests.
- Indirect administration questionnaires: in this case,the interviewer records the answers provided by the subject. The interviewer is therefore necessarily present.

Indirect administration gives the best results, but requires greater resources; when the number of people to be interviewed is very large or the wording extremely precise, direct administration should be used.

[1] Surveys are one of the most affordable ways to collect quantitative data.

[2] This is principle number 12 of the European Statistics Code of Practice, which is intended to ensure that statistics accurately and reliably describe reality.

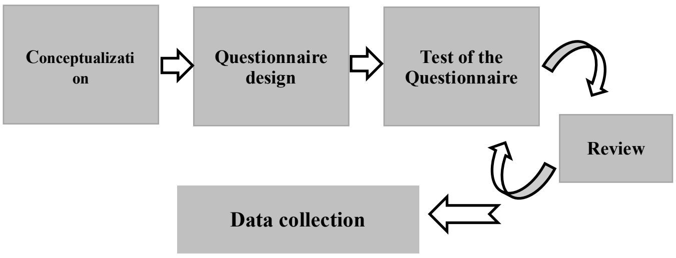
International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

The questionnaires can be administered in three ways:

- *By mail*: questionnaires are mailed to individuals in the selected sample.
- *By interview*: the questionnaire is completed during or after a face-to-face meeting between an interviewer and the respondent.
- *By telephone*: only a very short questionnaire can be submitted.

The risk in the first method is that there may be a greater or lesser percentage of non-responses, which distorts the representativeness of the sample. In contrast to interviews, the composition of the sample is easier to respect in this case, but the possibility of refusal or absence of respondents should not be forgotten.
When developing a questionnaire, the whole cycle of questionnaire design and testing should be covered. Five main steps should be distinguished (see Figure 2).

Figure 2: The five stages to be covered by a questionnaire



a. Conceptualisation phase:

This is the most important phase that comes before we even start thinking about the formulation of the questions, the conceptual basis of the questionnaire must be operationalised from the literature review.

The complexity of theoretical concepts entirely requires a strict selection of empirical features (indicators) that can be observed in a survey. These indicators are considered an appropriate representation of the concept. On the other hand, the development of the conceptual framework is obviously more important for completely new surveys, whereas in existing surveys the concepts are well established. In this case, less attention is required at this level.

The main output of this step is the list of target variables (in addition to basic decisions on the target population, data collection mode, etc.).

b. Questionnaire design phase

The questionnaire design phase starts after the specification of the conceptual basis and the determination of the data collection method. Based on the content of the questionnaire and the requirements of the data collection method, the sequence of the thematic sections of the questionnaire is decided.
The form and choice of answers proposed will be specified according to three criteria:

- Clarity[3] (understanding);
- Neutrality[4] (authenticity of answers)
- Adequac [5] which refers to the ability of the interviewees to answer the questionnaire. Three elements are important at this level: the length of the questionnaire, the order and the orientation of the questions.

Furthermore, the strategy for designing a questionnaire is relatively simple. It involves ensuring four important things:

- That respondents understand the questions.
- That they are able to respond.
- Let them agree to respond.
- That the answer is formulated authentically and not influenced

c.Questionnaire testing phase

The third phase consists of testing the questionnaire on a representative sample of the panel in order to check the order of the questions and their comprehension, and then to correct the questionnaire if necessary in the light of any problems encountered. The objective of this test is to evaluate the ease of understanding, the degree of acceptance and the ease of interpretation. The test is therefore an absolutely necessary step that must be carried out with rigour.
In this phase two things need to be checked:

- That the answers to each question are consistent with the expected outcome;
- This can be done by analysing the testers' answers, or by discussing them better once they have answered.

As making a good questionnaire is much more difficult than you might think, this step is essential, even for the most experienced, to ensure that there are no ambiguities that you might have missed.

d. Review phase

Generally, two or more rounds of questionnaire testing are recommended. If a questionnaire has been modified as a result of the test results, a new round of testing is normally required. This involves testing the questionnaire at an early stage of its development, revising the questionnaire according to the test results, and then testing the revised questionnaire. This process may be repeated for two, three or even more test phases.

---

[3] The set of people to be interviewed, the 'sample', is drawn from a larger population, the 'parent population' (also called the 'reference population' or 'parent population').

[4] The information to be collected must be requested (via the questionnaire) in an objective manner. This means ensuring the authenticity of the answers. A questionnaire is considered neutral when it prejudges the possible answers as little as possible.

[5] This criterion therefore refers to the fit of the questions to the characteristics of the questions to the characteristics of the respondents.

International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

Different methods of testing the questionnaire can be used during each phase of the test. In ongoing surveys, the evaluation of previous waves of the survey can make an important contribution to the revision of the questionnaire.

C. Choice of Measurement Scale

The method most adapted by researchers in management science is the scale of measurement which consists of expressing the intensity of its approval and adheres directly to the respondents, proposals to be judged according to a predetermined scale, it includes a statement expressing the state favorable or unfavorable to the object of interest. Each answer reflects, therefore, a degree of satisfaction with the existence of the elements of the theoretical model. A component of the scale is called a modality. The number of modalities can naturally vary, but generally there are 5 or more.
The score for each type of response is calculated by adding up the propositions that are favourable to the dimension being measured. However, this method implies assuming that each item has a weight similar to the average weight.
A pre-test generally ensures that the items are consistent with the variable to be measured. However, it is necessary to supplement this reflection with statistical (factorial) analyses in order to verify the dimensionality of the measurement scale (coherence) and to remove items relating to outliers.
Therefore, we adopted the five-point scales (ranging from 1 to 5 according to the following table), to be used in quantitative studies to achieve consistent measures and to seek a fit between the abstract ideas used to understand the social world, and what happens in the real (empirical) world (W. L. Neuman & Kreuger, 2003). The advantage of such a design lies in its ability to make a mental map of the respondent's evaluation obvious to the researcher (Schindler & Cooper, 2001).

Table 1: Five-point Likert Scale

Likert scale	**Scale coding**
Not at all in agreement	1
No agreement	2
Neither agree nor disagree	3
I agree.	4
Totally agree	5

D. Other forms of items

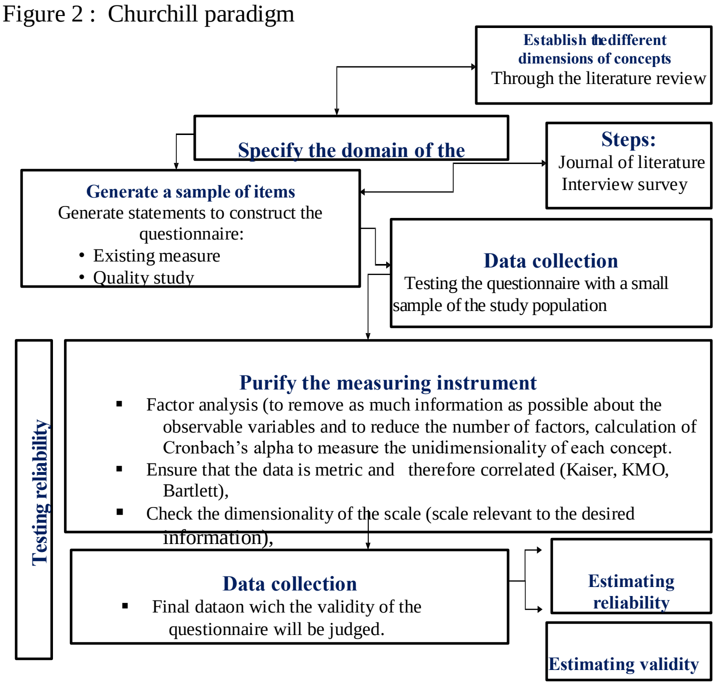
Generally, a question can be defined as an interrogative sentence that seeks an answer to assist in information gathering, testing and research. Good questions produce accurate answers and help to collect usable quantitative and qualitative data.
Historically, questions have evolved well into other types of questions for gathering information. Although the types of questions used in the design of a research study, these are determined by the information required, the nature of the study, the time needed to answer them and the budgetary constraints of a study. Thus the quality of the questions, i.e. the art of asking the questions, allows for in-depth knowledge to be gained, informed decisions to be made and effective solutions to be developed.
There are several types of questions. The most frequently answered are the dichotomous questions. The latter is usually a closed "Yes/No" question and is used for basic validation. In the context of this thesis work, it was essential to focus on these types of questions, for example this question is used to find out the gender of the respondent, this can allow us to group the data into two groups in order to test moderation for example.
Another type of question is proposed. This type refers to multiple-choice questions. These questions are a type of question in which respondents are asked to select one (single selection multiple-choice question) or more answers (multiple selection multiple-choice question) from a given list of options. The multiple-choice question consists of either a single correct answer, one or more correct answers, incorrect answers, etc. For example, the questionnaire has several single-selection multiple-choice questions such as "What is your highest degree? Here the respondent is asked to provide only one answer from a list of six proposals.

III .METHODOLOGY FOR THE EXPLORATORY EVALUATION OF MEASUREMENT INSTRUMENTS

Based on the methodology presented by the Churchill paradigm, the construction of multi-scale questionnaire-type measurement instruments is based on two distinct but complementary phases:

- An exploratory analysis to explore the measurement instruments (this is what we will see in this work).
- A confirmatory analysis to confirm or refute the research hypotheses.

Figure 2 : Churchill paradigm



The idea is to first test the questionnaire with a small starting sample. The test consists of purifying the measurement instruments by eliminating items that do not contribute

International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

significantly to the measurement of indicators measuring the concepts.
The exploration therefore consists of :

- Focus on the literature review to identify and specify the domain of the construct.
- The interviews and practical examples enable the researcher to generate indicators for measuring the concept.
- The collection of the first data is used to purify the measurement instrument in order to develop a second version of the questionnaire.

Figure 1: Protocol for the exploratory examination of instruments

Reliability of the measuring instrument
↓
Evaluation of the measuring instrument
↓
Factor structuring and scale cleaning

In the following paragraphs, it will be interesting to examine the dimensionality and factoring of the measurement scales by applying the principal components method without and with rotation. The objective of this step in the analysis is to further refine the measurement instruments and to reconstruct the variables in such a way that orthogonal relationships between the factors selected for the four variables of this dimension are known. The tests are performed with the *Stata* software package.

A. Reliability analysis

Reliability analysis is used to characterise measurement scales made up of various items - in the case of a questionnaire, questions. The procedure used calculates several measures that allow the reliability of the scale to be assessed and also provides information on the relationships between the different items. In our research, we often speak of internal reliability for measures with several items.
The reliability of a measuring instrument depends on the randomness of the measurement error (Evrard et al., 2003). The lower the randomness, the more reliable the instrument is. To assess the reliability of the measurement scales we have mobilised in this doctoral research, we will use the most widespread indicator in management science research.

Cronbach's alpha coefficient (1951), known as the internal consistency coefficient (measures the correlation between the scale and its various items). It is calculated on the basis of the average variances and covariances of the items, according to the following formula

$$\rho^2(y,t) = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k}\sigma^2 y_i}{\sigma^2 y}\right] \qquad \text{(Eq.1)}$$

With : $\boldsymbol{yi}$ For each i represent the scores of $\boldsymbol{k}$: Items of the questionnaire. $\boldsymbol{\sigma 2}$: is the variance of the sum of the items equal to the sum of the elements of the variance covariance matrix.
The Cronbach's alpha coefficient[6] can take on several values, from 0 to 1[7]. This index reflects a degree of homogeneity that is all the higher the closer the value is to one. Conversely, a Cronbach's close to 0 indicates that the measurement scale is unreliable. Al though there is no consensus in the literature on research methodology, many authors consider that a Cronbach's value of 0.6 [8] is acceptable for exploratory research (Malhotra, 2004; Mak, 1989; Evrard et al., 1993; Usunier et al., 1993, Peterson, 1995). Evrard et al (2003) consider that an alpha value between 0.6 and 0.8 is acceptable in an exploratory study. On the other hand, Nunnally (1967) and Peter (1979), reach more or less the same conclusion by setting the acceptance interval between 0.50 and 0.60.

However, it is a measure of the internal consistency of the questionnaire. In the first place, a high coefficient can then be interpreted as a good overall consistency of the items within the questionnaire. But a good general coherence does not mean that all the items, taken individually, are coherent with the others. It will be interesting to know the individual consequence of each item, in practice this step is primordial since it is important to know the effect of each item in the group of items constituting the construct or the dimension measured.

Table 3: Decision rule for the reliability of a measurement scale

Value of α	$\boldsymbol{\alpha > 0,8}$	**$0.6<\alpha<0,8$**	$\boldsymbol{0,5 < \alpha < 0,6}$	$\boldsymbol{\alpha < 0,5}$
Internal consistence	High	Average	Low	Very low
Decision		***Acceptable***	acceptable in the case of an exploratory study	***Unacceptable***

Various software and packages facilitate the calculations. In particular, STATA, SPSS, R, etc. offer a fairly simple handling of the estimation of *Cronbach*'s alpha, and an easy examination of inter-item correlations and covariances.

B. Analysis of factoring and dimensionality

[6] This coefficient can be interpreted as an estimate of the reliability coefficient of the questionnaire.
[7] In practice, the closer Cronbach's alpha is to 1, the more correlated the items are with each other and the more internally consistent the scale is. In this case, the items can be added together to form an overall score for this scale, as they are supposed to measure the same phenomenon.
[8] Cronbach's Alpha up to levels of 0.7 - 0.8 indicating good internal consistency of the scale and the range 0.8 - 0.9 reflecting excellent consistency between items.

There are many types of factor analysis, but principal component analysis[9] is perhaps the simplest and most widely used. Principal components are synonymous with factors, and generally the components in principal component analysis are often components are often referred to as factors. We will use the two terms interchangeably here. interchangeably here.
Principal component analysis (PCA) is therefore a multivariate technique known as interdependence analysis, which consists of rotating the axes of the factors around the point of origin in order to redistribute the variance to be explained more equitably. This is one of the most commonly used factorial methods. It is particularly suitable for quantitative, continuous variables that are a priori correlated with each other. Once the data are projected into different planes, the proximities between variables are interpreted in terms of correlations, while the proximities between individuals are interpreted in terms of the overall similarity of the observed values.

a. Scaling: PCA factoring without rotation

smaller factors we should ignore, as they explain the least amount of total variance. One of the main criteria used is the *Kaiser* or *Kaiser Guttman* criterion, which consists of ignoring factors whose eigenvalues are equal to or less than one. Since the maximum amount of variance that can be explained by a variable is **one**, these factors are really only equivalent to the variance of a variable.
The first factor will always explain the largest proportion of the overall variance, the second second factor will explain the second largest proportion of variance not explained by the first factor, and so on. factor, etc., with the last factor explaining the smallest proportion of the overall variance. variance. Each variable is correlated with each factor. As the first factor explains the largest proportion of the overall variance, the correlations or loadings of the variables will, on average, be the highest. will, on average, be highest for the first factor, lowest for the second factor, and so on.

In practice, the following steps should be followed:

- PCA without rotation is applied to all the items forming the variable under study.
- We detect the values of "*Eigen value"* which are higher than 1[10] . And make sure by a simulation that its values have not been modified. To decide on the number of principal components with p components or p−1 principal components so that the unexplained variability does not increase.
- Retain the principal components for the latent variable.

b. *Dimensionality of measurement scales: PCA with rotation*

The initial (retained) principal components that explain most of the variance in the measurement variables are of the measurement variables are rotated to clarify their significance. The rotation of factors can be done in different ways. We will discuss two of these approaches. of them. The most common form of rotation is called VARIMAX, in which the factors are not related or orthogonal to each other, i.e. the factors are are not related or orthogonal to each other, i.e. the scores of one factor are not correlated with the scores of the other factors. Are not correlated with the scores of the other factors. The VARIMAX rotation attempts to maximise the variance explained by the factors by increasing the correlation of variables that are highly correlated with them and decreasing the correlation of variables that are not. correlated with them and decreasing the correlation of variables that are weakly correlated with them.

The main objective of this method is to obtain a clear structure of factor weights, i.e. factors that are clearly marked by strong correlations with some variables and weak correlations with other variables. In an analysis with many variables, it is necessary to rank the items in terms of decreasing size for each factor in order to see more clearly which variables contribute most strongly to each factor.
In addition, the value of the variance explained by the two rotated VARIMAX factors is the sum or eigenvalue of the squared loadings for each factor, divided by the number of variables. These proportions are naturally different from those of the original principal components without rotation because of the change in the weights of the variables in relation to these factors.

In addition, the value of the variance explained by the two rotated VARIMAX factors is the sum or eigenvalue of the squared loadings for each factor, divided by the number of variables. These proportions are naturally different from those of the original principal components without rotation because of the change in the weights of the variables in relation to these factors.

Ultimately, the following steps should be followed:

- Perform a Principal Component Analysis with rotation: VARIMAX with Kaiser normalisation. (Orthogonal rotation).
- Deduce the factor structure of the latent variable according to the selected components.

C. Structural evaluations of measurement scales

Structural evaluation of measurement scales where the KMO[11] (Kaiser-Meyer-Olkin) test is used. This is a factorial solution adequacy index that measures the inter-correlation of items. The closer the value is to 1, the more *factorizable* the items are, and PCA (Principal Component Analysis) can therefore be applied.
This measure gives an overall picture of the quality of inter-item correlations. The KMO index11 varies between 0 and 1

[9] PCA for principal component analysis
[10] According to the Kaiser rule, only factors with an eigenvalue greater than 1 are retained.
[11] The KMO index is sometimes referred to as the MSA "Measure of Sampling Adequacy" in Anglo-Saxon software. But the French translation "mesure d'adéquation de l'échantillon"

International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

and gives additional information to the examination of the correlation matrix. Its interpretation is as follows:

Table 3: Interpretation rule for the KMO test

Valeur du test KMO	< 0,5	[0,5 ; 0,6]	[0,6 ; 0,7]	[0,7 ; 0,8]	[0,8 ;0,9]	> 0,9
Interprétation	Inacceptable	Insuffisient	Médiocre	Excellent	Méritoire	Excellent

Next, the relationships between the original variables and the components are examined. The stronger this relationship, the more the variable is "explained" by the factor. This relationship, which is expressed by a number between -1 and +1, is called the factor *loading* of the variable on the factor. A variable is only considered to be associated with a factor if its factor loading exceeds **0.3** in absolute value.

In order to improve the structuring of the factors of each of the measurement instruments used, we will carry out a PCA with orthogonal rotation of the VARIMAX type with Kaiser normalisation, generally used when it comes to measurement scales for which independent components are expected (Evrard et al., 2003).

This method makes it possible to obtain a clear factorial structuring by determining for each factor its weight on the different factors (axes). The items are then assigned to the different factors according to their weight. For each factor, only the items with a strong contribution are retained, while ensuring the significance of the percentage of variance explained and Cronbach's alpha coefficient.

IV. APPLICATION

We consider a questionnaire with several questions (items) grouped into theoretical variables. Our application concerns a latent variable named "*Decision evaluations of the line manager"*. And it is measured by six separate questions and structured according to a five-point *Likert* scheme as shown in the following table.

Table 2: Measurement scale for the latent variable "Decision evaluations".

Question/ Secondary issues	Scale of measurement				
Your decisions ...	*Not at all in agreement*	*No agreement*	*Neither disagree nor agree*	*I agree*	*Totally agree*
Are taken ***routinely*** (Q70)	☐	☐	☐	☐	☐
Are taken ***repeatedly*** (Q71)	☐	☐	☐	☐	☐
Are taken ***without involving*** my colleagues or superiors (Q72)	☐	☐	☐	☐	☐
Could be ***automated*** (Q73)	☐	☐	☐	☐	☐
Need ***experience*** and hindsight (Q74)	☐	☐	☐	☐	☐
Are so ***complex*** that they require computer use (Q75)	☐	☐	☐	☐	☐

As a reminder, the approach used for cleaning is the reduction of the random error of the measurement of the concepts studied (Examination of the internal consistency of the items). It is achieved by applying the following two steps: The first step is therefore to :

- ✓ Calculate the Cronbach's alpha coefficient. It is the usual measure of the internal consistency of a set of measurement indicators. It allows to estimate the extent to which an item can weaken the internal consistency[12] of a multiple scale.
- ✓ Remove items that weaken the Cronbach's alpha value below 0.60.

A. Analysis of internal reliability for the variable "Evaluation of decisions".

The objective of this application is to determine the reliability of the measurement instrument we have used to conduct this research. In our case, we use so-called Likert measurement scales.

The aim of this test is to improve its overall quality in order to be able to select *the* "*best*" items that reflect the constructs under study.

```
. alpha  Q70 Q71 Q72 Q73 Q74 Q75 , item

Test scale = mean(unstandardized items)
```

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
Q70	42	+	0.6369	0.4252	.2567364	0.5515
Q71	42	+	0.5815	0.3870	.2848432	0.5703
Q72	42	+	0.4888	0.1828	.3313589	0.6559
Q73	42	+	0.7402	0.5318	.2009872	0.4969
Q74	42	+	0.3929	0.1842	.3553426	0.6328
Q75	42	+	0.6742	0.4496	.2360046	0.5380
Test scale					.2775455	0.6229

The reliability analysis applied to the items of this variable shows that five items are retained instead of six. The value of Cronbach's alpha in this case indicates a value that is close to 0.632 after eliminating item number 74 (Q74), which affirms that the scale presents a satisfactory internal consistency for the constructs.

```
. alpha  Q70 Q71 Q72 Q73  Q75 , item

Test scale = mean(unstandardized items)
```

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
Q70	42	+	0.6393	0.4101	.3547232	0.5687
Q71	42	+	0.5809	0.3701	.3975029	0.5901
Q72	42	+	0.5423	0.2242	.4343786	0.6660
Q73	42	+	0.7707	0.5602	.2496129	0.4808
Q75	42	+	0.6561	0.4036	.3404955	0.5703
Test scale					.3553426	0.6328

[12] The most frequently used method of reliability estimation in psychometrics is the internal consistency method, also known as homogeneity, unidimensionality and internal consistency.

International Journal on Optimization and Applications
IJOA. Vol. 2, Issue No. 2, Year 2022, www.usms.ac.ma/ijoa

B. Application of the CPA

The evaluation of decisions is a variable defined from two items following the reliability analysis. First, it is interesting to apply a principal component analysis without rotation to the variable "Decision evaluation". This analysis will allow us to identify a clear factor structure with reliable internal consistency.

```
. pca    Q70 Q71 Q72 Q73 Q75

Principal components/correlation                 Number of obs    =         42
                                                 Number of comp.  =          5
                                                 Trace            =          5
    Rotation: (unrotated = principal)            Rho              =     1.0000
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.106	.778279	0.4212	0.4212
Comp2	1.32772	.586856	0.2655	0.6867
Comp3	.740861	.280383	0.1482	0.8349
Comp4	.460478	.0955314	0.0921	0.9270
Comp5	.364946	.	0.0730	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Unexplained
Q70	0.5010	-0.3157	0.4471	0.5310	-0.4092	0
Q71	0.4895	-0.4673	0.0565	-0.2577	0.6873	0
Q72	0.2419	0.7248	0.1696	0.4163	0.4626	0
Q73	0.4970	0.3942	0.1884	-0.6630	-0.3500	0
Q75	0.4514	0.0347	-0.8559	0.1969	-0.1538	0

Indeed, principal component analysis without rotation allowed us to retain two main factors (*Engen-value* values greater than one). The two principal components capture about 68.67% of the overall variability.
The first factor alone captures 36.70% and has a trace value of 1.83. The second factor absorbs 31.97% of the observed variance with a trace value of around 1.59.

The structure retained for the "decision evaluation" variable is a two-dimensional structure. Indeed, principal component analysis without rotation allowed us to retain two main factors (*Engen-value* values greater than one). The two principal components capture approximately 68.67% of the overall variability.
The principal component is exploratory factor analysis is to arrive at a parsimonious conceptualisation of latent traits, by determining the number and nature of a small set of factors explaining the correlation networks among a set of variables (Fabrigar et al, 1999). This is based on what Pohlmann (2004) considers to be the fundamental theorem of factor analysis that the correlation between two variables depends on the similarity of their relationship with the latent factors.

This technique, which aims to synthesise the information, consists firstly of verifying that the dimensions of the construct are clearly identified by common factors that are well distinguished and meet the criteria of convergent and discriminant validity (explaining the variability of the latent variable). And secondly to carry out a principal component analysis without rotation in order to determine the number of factors of the construct. The number of components is equal to the number of items.
After factoring, the results show that this factor is composed of all the items retained from the previous analysis (reliability of the measurement scales). Indeed, after rotation, items Q70, Q71, Q72, Q73 and Q75 are retained for better decision making. In addition, this analysis allowed us to retain a dimensional distribution of these items according to the estimated factors. In particular, items Q70, Q71, Q75 for the first factor and Q72, Q73 for the second factor.

```
. rotate, kaiser blanks(0.3)

Principal components/correlation                 Number of obs    =         42
                                                 Number of comp.  =          2
                                                 Trace            =          6
    Rotation: orthogonal varimax (Kaiser on)     Rho              =     0.5952
```

Component	Variance	Difference	Proportion	Cumulative
Comp1	1.9529	.334718	0.3255	0.3255
Comp2	1.61818	.	0.2697	0.5952

Rotated components (blanks are abs(loading)<.3)

Variable	Comp1	Comp2	Unexplained
Q70	0.5256		.4124
Q71	0.6012		.3084
Q72		0.7192	.1797
Q73		0.6219	.2692
Q74	0.3876		.719
Q75	0.3674		.5403

```
. pca   Q70 Q71 Q72 Q73 Q75 , component(2)

Principal components/correlation                 Number of obs    =         42
                                                 Number of comp.  =          2
                                                 Trace            =          5
    Rotation: (unrotated = principal)            Rho              =     0.6867
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.106	.778279	0.4212	0.4212
Comp2	1.32772	.586856	0.2655	0.6867
Comp3	.740861	.280383	0.1482	0.8349
Comp4	.460478	.0955314	0.0921	0.9270
Comp5	.364946	.	0.0730	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Unexplained
Q70	0.5010	-0.3157	.3391
Q71	0.4895	-0.4673	.2053
Q72	0.2419	0.7248	.1792
Q73	0.4970	0.3942	.2734
Q75	0.4514	0.0347	.5692

C. Overall assessment: KMO test

The output above shows that the goodness-of-fit index for this factorial solution is 0.59. This value seems poor when referring to the acceptance thresholds of the KMO test. This value seems poor when referring to the acceptance thresholds of the KMO test. However, it seems that the set of items selected can constitute a coherent double structure allowing

```
. estat kmo

Kaiser-Meyer-Olkin measure of sampling adequacy
```

Variable	kmo
Q70	0.6404
Q71	0.5522
Q72	0.4452
Q73	0.6088
Q75	0.7513
Overall	0.5897

to approach the variable "Evaluation of decisions".

V. Conclusion

This methodological work par excellence outlines the different stages required to carry out an exploratory factor analysis, the importance of which lies in the choice of techniques and methodology to be adapted. Indeed, it is important to present the research tool that is based on a questionnaire survey. In developing the questionnaire, the whole cycle of designing and testing the questionnaire (pre-test) was covered. The questionnaire went through several stages including the questionnaire design phase which started after the specification of the conceptual basis and the determination of the data collection mode. Thus, the factor analysis approach allows the identification of latent factors from directly measured variables. That is, it defines each latent variable (construct) by associating a number of measured variables (item) with it.

The exploratory study involved another analysis that falls within the framework of factorial component analysis, which aims to find the factors that summarise the data or their characteristics. The approach consists of conducting the three main validity stages most commonly used: PCA without rotation, PCA with varimax rotation and the KMO test. At this level, the questionnaire is considered purified and valid for a second administration on a larger sample.

REFERENCES

[1] A. David. (2000), Logique épistémologie et méthodologie en sciences de gestion : trois hypothèses revisitées, David, A., Hatchuel. and Laufer, R. Les nouvelles fondations des sciences de gestion, Paris, Vuibert, p. 83-109.

[2] F. Pearson. (2012), Méthodologie de la recherche en sciences de gestion-Marie-Laure Gavard-Perret, David, G. Christophe, H. and Alain Joliber, Méthodologie de recherche, Raymond-Alain Thietart et al, 4th edition, DUNOD.

[3] J. Barbet. (1988), Faut-il encore parler d'approche qualitative et d'approche quantitative, Recherche et Applications En Marketing, 3(1), 75-89.

[4] J. Cronbach. Coefficient *alpha and the internal structure of tests*, Psychometrika, Springer Science+Business Media, vol. 16, September 1951, pp. 297-334 (ISSN 0033-3123 and 1860-0980, DOI 10.1007/BF023105559.

[5] J. Hair., R. Anderson., Tatham, R. and Black, W. (1998), Multivariate data analysis (5 th ed), New York: Prentice Hall.

[6] L. J, Cronbach. (1951), Coefficient alpha and the internal structure of tests, Psychometrika, 16, 297-334.

[7] J. R, Busemeyer., L. E, Jones. (1983), Analysis of multiplicative combination rules when the causal variables are measured with error, Psychological Bulletin, 93, 549-563.

[8] L. V, Campenhoudt., R, Quivy. (2011), 1st step: the starting question, in Handbook of Social Science Research (Vol. 4, pp. 24-37).

[9] P. Roussel. (2005), Méthodes de développement d'échelles pour questionnaires d'enquête, Management des ressources humaines, méthodes de recherches en sciences humaines et sociales, edited by Roussel, P., Wacheux, F., Brussels.

[10] R. Cudeck. (2001), Cronbach's Alpha on two-item scales, Journal of Consumer Psychology, 10 (1/2), 55-69.

[11] T. Jolliffe. (1986). Principal component analysis. Springer-verlag.

[12] T. W, Anderson. (1963). Asymptotic theory for principal component analysis. Annals of Mathematical Statistics 34, 122-148 - Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 417-441.