Satisfaction of healthcare consumers

Analysis and comparison of different methodologies

Inês de Carvalho Vieira

Universidade de Lisboa – Instituto Superior Técnico

December 2020

Abstract

In the current world, patient satisfaction and the factors influencing it are becoming one of the principal concerns, not only for health organisations but also for patients themselves. Assuring the quality of the provided services is essential for the fulfilment of patients' expectations and needs. Thus, with a constantly evolving health system, factors that need improvement keep changing over time. That is the main focus of patient satisfaction, to point out factors and areas that have to be reviewed in order to provide the best possible service.

This study aims at evaluating patients' satisfaction regarding a secondary healthcare unit, in particular, the internment service. The main purpose is to identify which factors influence satisfaction, along with the satisfaction levels perceived by patients. To accomplish this, a methodological comparison (using factor analysis, structural equation modeling, ordinal logistic regression, and multicriteria satisfaction analysis) is performed with the goal to contrast the different results that may arise from the usage of different methodologies. After implementing the methods, it is concluded that, out of the eleven analysed factors, seven influence satisfaction, being them: accommodations, auxiliary staff, exams and treatments, medical staff, food quality, volunteering staff, and obtained information. It is expected that with this information, hospital managers are able to allocate the available resources in a more efficient manner, improving the healthcare experiences of Portuguese patients.

Keywords: Patient satisfaction; Quality of healthcare; Factor Analysis; Structural Equation Modeling; Ordinal Logistic Regression; Multicriteria Satisfaction Analysis;

1. Introduction

Patient satisfaction is a concept that has long been neglected and cast aside but is becoming gradually more important. Inclusively, Donabedian (1988) includes it as an outcome of healthcare services; hence, it is of extreme importance to evaluate satisfaction with care. Measuring healthcare quality and satisfaction does not only constitute an indispensable element for appropriate management of resources but also allows focusing on its users' preferences, giving them a chance to participate in the construction of a customized health service, better fitted to their needs and expectations (Abrantes, 2012). With this information, managers can more efficiently allocate resources to improve patients' experience and satisfaction (Al-Abri and Al-Balushi, 2014). When talking about public hospitals, there may not be a financial interest in performing these studies since they are not particularly interested in profit. However, with the increase of market competitivity, for private companies, it is vital to meet patients' needs, satisfying them, so they become loyal to the organization (Lovelock and Wright, 2001). The purpose of health care, whether provided by a public or private entity, is to guarantee the best service level. Nevertheless, due to resources' restrictions, it is not always possible to provide the best care in every service dimension. That is why the present study can be a useful tool for health care improvement. The evaluation of which dimensions influence patient satisfaction provides valuable insight on how to allocate available resource in the best way possible. Despite the existence of a strong legal and political commitment to the well-being of society, health inequalities are an issue in Portugal. Hence, adjustments need to be made to increase the efficiency and quality of health services. In Portugal, specifically, satisfaction studies are held on a national level. This keeps the results from translating the actual reality of each health unit. Therefore, the present study focuses on evaluating the satisfaction of Portuguese patients with reference to a secondary healthcare unit. For this end, an inquiry is distributed, and patients are asked to measure their satisfaction towards that particular service.

2. Satisfaction in health

2.1. Definition of satisfaction

Over the past few decades, given the customer-orientation philosophy and continuous improvement principles of modern enterprises, consumer satisfaction has gained recognition as a measure for quality in many public sector services (Ferreira, 2019). Customer satisfaction had its origins in the decade of 1960 but was only given the proper attention in the '80s. At this time, there was still a lot to be unveiled about this concept. Cardozo (1965) stated that customer satisfaction with a product is influenced by the effort spent to acquire it and the expectations concerning it. With the turn of the century, more complex conceptualizations emerged. For Evans et al. (2006), satisfaction/dissatisfaction is more than how well a product or service performs because it also involves the consumers' attitude and feelings. For instance, anticipatory satisfaction may not be related to the actual performance of a product/service but rather to the consumer's imagined ideas about how the product is going to function.

2.2. Patient satisfaction & quality in healthcare

The incessant demand for improved results and quality of health services offered is of extreme importance in the development of a more effective organizational policy, adjusted to the needs of the patients. Health organizations recognize that service quality is especially pertinent regarding promotion and public image of the healthcare market (Qin and Prybutok, 2013). Patient satisfaction and quality of health services are a priority for the services industry due to increasing consumption and are critical elements for the long-term success of health institutions (Ramsaran, 2005; Kaya et al., 2020). Even though satisfaction is an essential aspect of quality, the relationship between these two concepts is not linear. On the one hand, the results of satisfaction studies can be ambiguous and may not always be impartial. Given that patients evaluate physician's performance, for which most of them lack the necessary abilities, results can be based on affinity and not on the health professional technical skills. On the other hand, providers may have to face a trade-off between providing satisfaction to their patients or better treatment outcomes (Ferreira et al., 2018). One of the leading critics of patient satisfaction ratings is the incapacity to rationalize medical care expectations, that can be affected by previous healthcare experiences (Schoenfelder et al., 2011). The same happens with the other two components. The patient will be satisfied with the process if symptoms are reduced, and the outcome will be favourable if there is a recovery, demonstrating that perception of received care has met the prior expectations. Throughout his framework, Donabedian regarded "outcome" as the most crucial aspect, defining it as a change in a patient's current and future health status that can be confidently attributed to antecedent care (Ferreira et al., 2018).

3. Literature review

This research, an overview of an article submitted to an ISI peer-review journal that is awaiting revision, was performed respecting the guidelines of the Preferred Reporting Items for Systematic Reviews and MetaAnalysis (PRISMA) statement. A checklist of 27 parameters, including the title, abstract, methods, results, discussion, and funding, was taken into consideration to ensure complete reporting of systematic reviews. The PRISMA statement starts with the identification of possible studies to include in the revision, after searching in multiple databases. Papers were searched in the Scopus, Web of Science, and PubMed databases during June 2020. After testing several keywords, the search strategy used the term "patient satisfaction" to extend the number of results. Reference lists from the collected articles were also searched for additional articles. Following such a statement, 153 studies met the inclusion criteria. Authors, country of study, sample, quality dimensions and drivers, methodology, dependent variable, and main factors affecting the satisfaction of each study were assessed to reach a conclusion. A dependent variable field was considered because, besides overall patient satisfaction, some articles studied proxies of satisfaction, such as willingness to recommend the hospital or willingness to return. In general, studies about patient satisfaction try to unveil factors associated with his/her overall satisfaction with one or more services (96% of the collected studies) or willingness to recommend the hospital/clinic (9%), instead. A smaller percentage of studies (7%) included both dependent variables.

3.1. Global analysis over the utilization and influence of satisfaction criteria and explanatory variables

The utilization and influence of each factor related to patient satisfaction were analysed. The percentage of utilization is the ratio between the number of studies using it and the total number of evaluated studies. The influence rate of a factor measures the relative number of papers concluding that this factor is critical for patient satisfaction. All factors related to patient satisfaction were analysed and clustered in terms of satisfaction criteria, and explanatory variables, regardless of the dependent variable used by researchers. Some past systematic reviews have revealed that interpersonal or social skills (like medical/nursing care and attitudes), technical skills, infrastructure and amenities, accommodations, environment, accessibility, continuity of care, and the outcome are the satisfaction criteria present in the majority studies related to satisfaction in healthcare.

3.2. Utilization analysis

The "quality dimensions and drivers" section was reviewed to analyse the most utilized factors. The fifteen most utilized factors were divided into criteria and explanatory variables researchers use to study patient satisfaction and may not correspond to the most important and influential factors of patient satisfaction. From the fifteen most used factors, eleven are criteria, and four are explanatory variables. On the one hand, doctor's characteristics, waiting time, medical care, and information provided have the highest utilization rates within the criteria. On the other hand, patient's social characteristics, patient's age, patient's education, and perceived health status also have the highest utilization rates within explanatory variables.

3.3. Influence analysis

This analysis resulted in fifty-six factors, divided into forty-seven criteria and nine explanatory variables. The three most influential criteria are medical care, waiting time, and communication with the patient. Despite not being on the top three, criteria related to doctor's social skills exhibit a high importance rate and should be noticed as well. It is interesting to note that researchers tend to conclude that criteria related to social skills of staff, such as communication, are more important than others, for instance, food quality and comfort. Also, criteria associated with the technical skills of staff appear to be less critical. It seems to be in line with some authors claiming that patients are usually unable to judge health professionals in those terms. Patient's age, perceived health status, and patient's education are the explanatory variables that researchers tend to consider as the most influential. The conclusions from previous studies saying that age, education, and self-reported health status have an evident and significant influence on the satisfaction outcomes were confirmed (Hekkert et al., 2009). Older patients or the ones with better self-perceived health status are typically more satisfied, while highly educated people are less satisfied with the healthcare services provided (Nguyen Thi et al., 2002; Rahmqvist and Bara, 2010).

Past reviews acknowledge interpersonal or social skills (like medical/nursing care and attitudes), technical skills, infrastructure and amenities, accommodations, environment, accessibility, continuity of care, and the outcome as the most important factors. In terms of explanatory variables, these reviews also point out the frequent use of variables like the patient's gender, age, education, and marital status. Despite the similarity of results between previous studies and this bibliographic review, some factors seem to occupy a place of relevancy not seen before. Waiting time and information provided are not present on previous reviews. On the one hand, waiting time is a determinant of dissatisfaction in healthcare, regardless of the stage in which the inpatient is. On the other hand, the criterion information provided may refer to any process of care since the patient enters the system until he/she leaves it.

4. Methodology

4.1. Methodology present in literature

The methodology implemented in the 153 studies was assessed and four main methods emerged. Regression analysis is the method that is used in most studies, with a utilization rate of 52%. It is important to note, however, that with the goal of simplifying this comparison analysis, six different types of regression analysis were clustered into one big group. Out of the 52% of studies that apply regression analysis, 31% use multivariate regression analysis, 26% employ OLR, 19% utilize linear regression, 11% use multiple regression analysis, 8% implement ordinary least squares regression, 3% use multilevel analysis, and 2% utilize stepwise regression. As it can be seen, there are many regression methods available and present in the literature. However, due to the characteristics of our sample, something that is further discussed, OLR is the method that will be used. Factor analysis comes in second place, with a 32% rate of utilization, followed by SEM (15%), and at last, MUSA (1%). From the 153 collected articles, 27 (18%) combined different methods in a complementary nature: factor analysis with regression analysis (16 of the 27 articles, or 59%), and factor analysis with SEM (11 of the 27 articles, or 41%). The difference in the level of utilization of each method can be due to the difficulty of implementation. SEM and MUSA are more complex than the other two, and thus harder to implement. Logistic regression and factor analysis are simpler and easier to implement, becoming more attractive to the researcher. Each of the four methods has different advantages and disadvantages that are explained in the following sections. These differences must be considered when pondering which method to use since different methods can deliver different results. For this specific project, the combination of the four methods is used both on a complementary and comparative nature. Factor analysis is the first method to be applied to the entire sample. It is employed in a complementary nature to eliminate redundancy that might exist. Subsequently, SEM, OLR, and MUSA are implemented, and the results of each method are compared.

4.2. Factor Analysis

Factor analysis is a mathematical model that explains the correlation between a large set of variables in terms of a small number of underlying factors (Mardia et al.,1994), using procedures that summarize information included in a data matrix and replacing original variables by a small number of composite variables or factors. The goal of factor analysis is to diminish the dimensionality of the original space and to interpret the resulting space, covered by a reduced number of variables that dominate the previous ones (Rietveld and Van Hout, 1993). Two models differ in purpose and computation: principal components analysis (PCA) and common factor analysis (Fabrigar et al.,1999). PCA transforms original variables, that can be or not correlated, into a smaller set of non-correlated variables. In common factor analysis the main goal is to explain covariance structure amongst original variables through a hypothetical set of common factors (unobservable) (Vilares and Coelho, 2011). Based on the common factor model, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have been presented (Thurstone, 1947).

4.3. Structural Equation Modeling

SEM is a statistical methodology that takes hypothesis-testing to the multivariate analysis. Multivariate procedures commonly used in market research are essentially descriptive or exploratory, so hypothesis testing is difficult, if not impossible. SEM generally involves the specification of an underpinning linear regression type model (incorporating the structural relationships or equations between unobserved or latent variables) together with several observed or measured indicator variables (Byrne, 1994). SEM can estimate dependence relationships, represent unobserved concepts in these relationships, consider measurement errors in estimation, and define a model explaining an entire set of relationships (Xiong et al., 2014). The general SEM can be also known as Linear Structural Relationships, a linear model that establishes relations between variables. This model can be organised into two sub-models: the structural and the measurement model. The structural model shows the relationship between latent variables and is synchronously estimated with the measurement model. When a model only contains observable variables, the structural model is reduced to path analysis. The measurement model shows the relationship between observed and latent variables. Its goal is to illustrate how well the observable variables measure the latent variables. Measurement variables are assessed by CFA. This method determines relations between observed and latent variables and tests them in order to confirm the suggested structure (Marôco, 2014).

4.4. Ordinal Logistic Regression

As previously mentioned, there are many types of regression analyses, each presenting different characteristics and being best suited for different types of variables and samples. OLR, that can be referred to as the proportional odds model, is a special type of multinomial regression, which can be advantageous when the response variable is ordinal (Koletsi and Pandis, 2018). Since that is the case with our sample, OLR was the regression method chosen. OLR is a statistical method where one variable is explained or understood based on one or more variables. The variable that is being explained is called the dependent, or response variable. The other variables used to explain or predict the response are called independent variables. Linear regression is the standard or basic regression model in which the mean of the response is predicted or explained based on a single predictor. The basic model is easily extendable such that it becomes a multivariable linear model, that is, a linear regression having more than one predictor (Hilbe, 2017). In an OLR model, the outcome variable has more than two levels. It estimates the probability of being at or below a specific outcome level given a collection of explanatory variables (Xing Liu and Koirala, 2012).

4.5. Multicriteria Satisfaction Analysis

The MUSA model was developed to measure and analyse customer's satisfaction from a specific product or service, but the same principles can be used to measure global satisfaction of a group of individuals regarding a specific service or operation that they interact with (Muhtaseb et al.,2012). The basic principle of MUSA is the aggregation of individual judgments into a collective value function, assuming that customers' global satisfaction depends on a set of criteria representing service characteristic dimensions. This preference disaggregation methodology is implemented through an ordinal regression-based approach used for the assessment of a set of marginal satisfaction functions in such a way, that the global satisfaction criterion becomes consistent with customer's judgments (Drosos et al., 2015). The method finds an additive utility function representing the satisfaction level of a set of customers based on their expressed preferences collected in a satisfaction survey's data. Customers are asked to give a satisfaction level for a service or a product, as well as a marginal satisfaction level for each one of its characteristics (Grigoroudis and Siskos, 2003). It is implemented through the principles of the Patient Satisfaction Model (PSM), with an objective function as follows:

$$minimize\ z = \sum_{q=1}^{p} (\alpha_d^{(q)+} + \alpha_d^{(q)-}) \qquad (1)$$

In the objective function there is a minimization of the sum of the non-negative error variables, $\alpha_d^{(q)+}, \alpha_d^{(q)-}$, for all q = 1, . . . ,p, where P = {1,..., q,..., p} represents a set of patients whose satisfaction respecting a hospital is being assessed; each patient q ∈ {1,..., p}, characterizes the hospital according to a single level of each scale Ej, for j = 1,..., n and E (Grigoroudis and Siskos, 2002). This function focuses on minimizing the deviation between patients' overall and partial judgements.

4.5.1. MUSA and the Kano's model

The Kano's diagram specifies three types of relationships between the degree of customer satisfaction and the fulfilment of expectations (Kano et al., 1984; Wang and Ji, 2010):

- Must-be Attributes (critical or necessary): customers take must-be characteristics for granted. If these requirements are not sufficiently met, customers will be dissatisfied. However, their presence does not contribute to customer satisfaction.
- One-dimensional attributes (high value-added or low value-added): their fulfilment is positively and linearly related to the level of customer satisfaction. The higher the level of fulfilment, the higher the degree of customer satisfaction, and vice versa;
- Attractive attributes (highly or less): fulfilment of attractive attributes will lead to greater satisfaction. However, since customers are not expecting these requirements, they will not be dissatisfied in the case of absence. These requirements are seen as 'pleasant surprises'.

The categorization according to the Kano's model is accomplished by comparing the weights that satisfied and dissatisfied patients assign to each requirement. For the subdivision into refined Kano's model categories, the weights of each criterion/subcriterion and the weight's centroid are compared.

5. Case study: The experience of Portuguese citizens with a public hospital's service

A patient satisfaction survey composed of 65 questions was delivered, in 2018, to patients in the internment service of a secondary healthcare unit, that due to privacy reasons cannot be named. The survey is comprised of one question regarding the partial satisfaction for each of the 53 subcriteria, one question about the global satisfaction for each of the eleven criteria, and finally one question concerning the global satisfaction with the service provided. A total number of 251 responses were gathered from patients between 20 and 92 years old (with an average of 42 years old), both male (27%) and female (73%), out of six different medical internment specialities: paediatrics (63%), gastroenterology (14%), nephrology (8%), urology (6%), orthopaedics (5%) and internal medicine (4%). The survey followed the official template of the Portuguese NHS, allowing an understanding of patients' partial and global satisfaction, translated through a seven-point Likert-type scale, where 1 means very dissatisfied, and 7 means very satisfied. A seven-point scale is considered to have the most common number of response alternatives such as stated by Cox III (1980) 'If the number of alternative responses was to be established democratically, seven would probably be selected'. Going into further detail about the evaluated parameters, the eleven criteria present on the survey are: *obtained information, accommodation's quality, visits, food quality, medical staff, nursing staff, auxiliary staff, administrative staff,* *volunteering staff, exams and treatments, and discharge process*. Comparing these criteria to the results gathered from the utilization statistical analysis, it is possible to conclude that every criterion has also been included on the surveys of previous studies. This is an indicator that the parameters assessed in this case study are reliable and aligned with past researches. The percentage of patients that evaluated each subcriterion with a determined satisfaction level is calculated to identify the service dimensions where patients seem mostly satisfied or dissatisfied. It is important to note that since the scale used on the questionnaires is a seven-point Likert-scale, the expected level of satisfaction is based on the sum of the answers of the fifth, sixth, seventh levels. *Nursing staff* is the dimension that mostly satisfies patients. It is followed by *auxiliary staff* and *food quality*. When assessing the areas where patients are mostly dissatisfied, *volunteering staff* is the criterion that provides the worst service. *Obtained information* and *visits* also leave patients with a feeling of dissatisfaction.

6. Results discussion and implications

6.1. Factor analysis results

Factor analysis was performed on SPSS software through the principal components' method along with a varimax rotation. Bartlett's sphericity test demonstrated a p-value<0,001, meaning that the null hypothesis was rejected and factor analysis to extract components could proceed. The Guttman-Kaiser rule was followed, meaning that only factors with eigenvalues higher than one were selected. Communalities present high values, which means that the extracted components are a suited representation of each criterion. The rotated component matrix shows the rotated component loadings as well as the correlations between variables and said component. Results show a clear grouping of subcriteria into components, with meaningful loadings. According to the rotated component matrix, it is possible to conclude that subcriteria can be grouped into nine components, instead of the eleven originally presented on the questionnaire. The explained variance of the nine components, 82.558%, throughout the three stages of the analysis, proving there is no significant loss of information from the original variables. The new nine components are the following: obtained information, accommodation, visits, food, medical services (medical staff + discharge process), health staff (nursing staff + auxiliary staff), administrative staff, volunteering staff, exams and treatments. The adequacy of the analysis was evaluated by multiple coefficients (ANOVA, Cronbach's alpha, intra-class correlation coefficient, Pearson's correlation coefficient, Spearman's correlation coefficient, Kaiser–Meyer–Olkin, Mann–Whitney U test, and independent t-test) demonstrating good adequacy and consistency throughout all coefficients. Regarding the possibility of analysing the sample in two distinct groups, being the patients divided by gender, it is proven by ANOVA, Mann-Whitney U and independent t-test that there are no significant differences that justify the division of data. An alternative analysis with the initial structure of eleven criteria is also performed. In this analysis, the dataset did not undergo exploratory factor analysis and criteria are treated as observable variables, using patient's survey responses, as opposed to constructs. To simplify the reckoning of each analysis, analysis A is assigned to the database that resulted from exploratory factor analysis which contains nine criteria, and analysis B refers to the analysis of the original dataset that did not undergo exploratory factor analysis and includes patients' criteria judgements.

6.2. Structural Equations Modeling results

6.2.1. Analysis A

From the rotated component matrix, a SEM path diagram was created on Amos SPSS software. The initial path diagram includes nine latent variables and 54 observable variables. Firstly, the existence of outliers was evaluated through Mahalanobis distance. In total, fourteen observations presented high distance values (>50) as well as p1 and p2 <0,001. These observations were removed from the dataset. A stability analysis followed these modifications. However, the indices still demonstrated poor adjustment, and further alterations were necessary. Secondly, modification indices were assessed. As a general rule, modification indices with values superior to eleven should be appraised (Marôco, 2014). Subcriteria attention of auxiliary staff and kindness of auxiliary staff (G_{72} and G_{73}, respectively) showed modification indices of around 200. With such high modification indices, removal of the subcriteria was recommended. Since these subcriteria were grouped into component health staff, still composed of eight other subcriteria, reliability of the component was maintained after their elimination. Despite these adjustments, modification indices' values were still superior to what was desired, translating a need for changes in the model. Thus, covariances between measurement errors were established, according to their modification indices. When a third stability analysis was performed in Amos software, results still showed poor adjustment. From there, subcriteria information provided by medical staff regarding patient's health state and kindness of professionals from exams and treatments (G_{54} and G_{103}) were removed due to model discrepancies (inflating X2, and decreasing GFI, CFI, PGFI and PCFI). These subcriteria belonged to a large component and were redundant in face of all available information. After this, covariances were, once more, established according to modification indices. Covariances between measurement errors of subcriteria within the same component are understandable, as well as easy to interpret and to explain since redundancy can naturally be present in these cases. A final stability analysis was performed. Results demonstrate good adjustment throughout the majority coefficients (X^2, X^2/df, GFI, RMSEA, CFI, NFI, PGFI, PCFI, PNFI) indicating goodness of fit and adequacy of the model. The final path diagram, displayed in Figure 1, includes nine latent variables and 50 observable variables. It is concluded that four criteria influence patient satisfaction, given their positive standardized regression weights and statistically significant p-values. *Accommodations* is the component with the highest loading (loading = 0.329 and p<0.001), thus, being the one that most influences patient satisfaction. It is followed by *exams and treatments* (loading = 0.277 and p<0.001), *medical services* (loading = 0.202 and p<0.001), and *health staff* (loading = 0.192 and p=0.004).

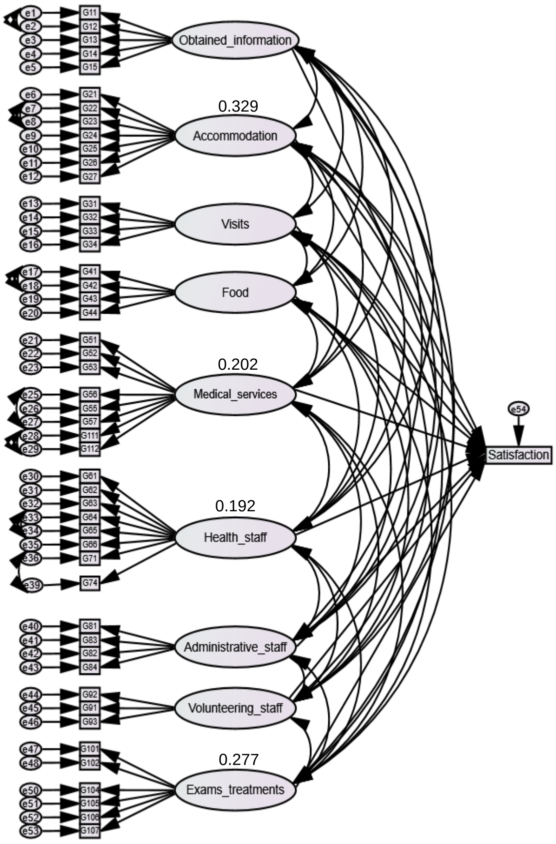

Figure 1. Complete path diagram (Analysis A). Source: SPSS software.

6.2.2. Analysis B

For this analysis, an alternative SEM model with 65 observable variables was designed on AMOS SPSS software. Since this model only contains observable variables, it is considered a path analysis model where an observable construct (endogenous variable) has a linear relationship with two or more observable indicators (exogenous variables). This type of relationship between variables is a characteristic of a formative model. As a direct consequence, removal of an indicator might alter the construct itself, therefore it is not recommended (Diamantopoulos, 1999). A problem related to this type of model is the measure of reliability and consistency. There is no universally valid measure to assess the reliability of formative variables (Coltman et al., 2008). Marôco (2014) states that the reliability of the model should be assessed based on the R^2 coefficient. Firstly, the existence of outliers is verified through *Mahalanobis distance*. In total, thirteen observations presented high distance values (>50) as well as p_1 and p_2 <0.001. These observations were removed, and the analysis proceeded with a total of 237 observations. Since no stability analysis can be performed in this model, R^2 is verified to assure the model is valid. The dependent variable *satisfaction* has a R^2 value of 0.645, which indicates that the model is adequate. Therefore, no further measures to increase reliability were explored. The final path diagram is presented in Figure 2 with the respective standardized regression weights. It is concluded that four criteria influence patient satisfaction. *Auxiliary staff* (loading = 0.408 and p<0.001) is the most influential criteria, followed by *exams and treatments* (loading = 0.395 and p<0.001), *medical staff* (loading = 0.362 and p<0.001), *and accommodations* (loading = 0.271 and p<0.001).

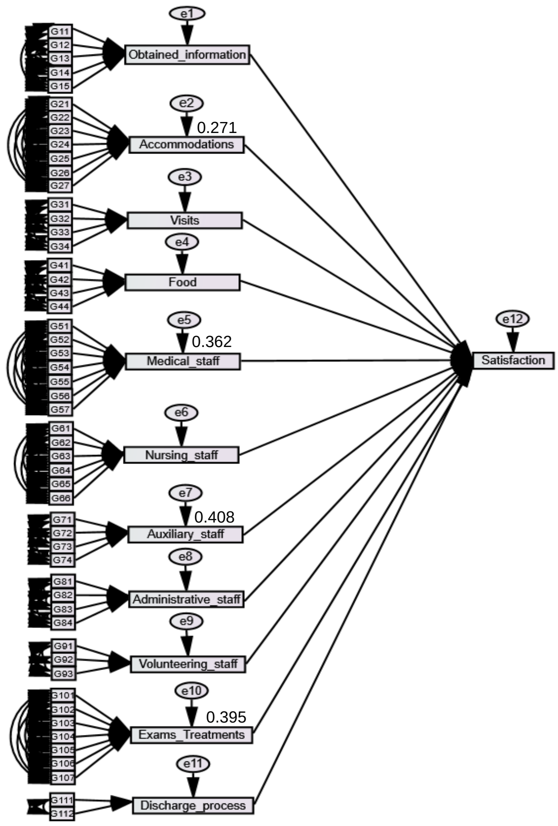

Figure 2. Complete path diagram (Analysis B). Source: SPSS software.

6.2.3. Results comparison

Since in analysis A criteria were treated as latent variables, their values were created by SEM and do not correspond to the real values used in analysis B. The differences that emerge might be due to misjudgements attributed, by SEM, to latent constructs. In conclusion, both analyses demonstrated rather similar results. Thus, criteria deemed as influential according to SEM, and common to the two analyses, are *accommodations, auxiliary staff, exams and treatments, and medical staff.*

6.3. Ordinal Logistic Regression results
6.3.1. Analysis A

From the likelihood ratio chi-square test, it is possible to conclude that the final model is a significant improvement in fit when compared to the null model with a significance level below 0.001. The Pearson and deviance tests show non-significant levels (1.000 each), indicating good model fit. Pseudo R-squared values were also evaluated and demonstrated good adjustment (>0.500). To finish this analysis, a test of parallel lines was performed, showing a non-significant result of 0.983, indicating that the proportional odds assumption is verified. Once the assessment of model fit measures is done, parameter estimates are estimated. *Accommodations* appears as the most influential predictor of patient satisfaction. The OR is exp(β)= 4.937, meaning that the odds of a patient being more satisfied increase by 4.937 for every unit increase on *accommodations*. *Exams and treatments* is the second most influential predictor of the dependent variable, with an OR of 2.673 and statistical significance (sig=0.000). *Medical service* is the last predictor of patient satisfaction, with an OR of 1.532 and a significance level of 0.012. The remaining criteria are not statistically significant (sig>0.005), thus, are not considered influential regardless of their log-odds and OR.

6.3.2. Analysis B

The likelihood ratio chi-square test, with a significance level of 0.000, shows that the final model is a significant improvement in fit when compared to the null model. The Pearson test is statistically significant (sig<0.005), demonstrating a bad model fit. The deviance test is non-significant (sig>0.050), indicating good model fit. Because of these differences, no conclusions regarding the model fit can be retrieved from these coefficients. Pseudo R-squared values were also evaluated and demonstrated good adjustment (>0.500). The final step of the adequacy analysis is the test of parallel lines that returned a non-significant value of 0.994, thus, the proportional odds assumption is verified. With the adequacy analysis finalised, the outputs of the ordinal logistic regression are estimated. *Auxiliary staff* is the criteria that most determines patient satisfaction. Exp(β) has a value of 3.582, meaning that for each unitary increase there is a predicted increase of 3.582 on the odds of *patient satisfaction*. E*xams and treatments* also influences *patient satisfaction*. With an OR of 2.646, for every unitary increase on *exams and treatments* there is a predicted increase of 2,646 on the odds of *patient satisfaction.* The last influential criterion is *accommodations*. There is a predicted increase of 2.401 on the odds of *patient satisfaction* for each unit increase on accommodations. Given the statistically non-significant levels of the remaining criteria, they are treated as non-influential, despite their log-odds and OR.

6.3.3. Results comparison

Once more, results from both analyses do not completely converge but can be seen as approximate, the justification given for the SEM method still applies. In lines with this, *accommodations, auxiliary staff,* and *exams and treatments* are the criteria seen as satisfaction predictors by OLR.

6.4. Multicriteria Satisfaction Analysis results

The results from the MUSA method, performed on MATLAB, are presented in Table 1. Along with the weights of every subcriterion and criterion, multiple indices are assessed to provide a deep insight into patients' preferences. Oppositely to what happened with SEM and OLR, for MUSA, results are not developed in two separate analysis (analysis A and analysis B). To put it simply, MUSA is only applied to the original database because this method cannot work with data returned by factor analysis. There is an exception to this rule, that is the usage of a categorical principal components analysis (CPCA), as seen in studies associated with satisfaction, such as Valle et al. (2011) and Vuković et al. (2012). According to MUSA, *food quality* is the most influential criterion. Since this entire set of attributes is composed of critical must-be requirements and given the high satisfaction indexes (not only of the criterion but also of the subcriteria), modifications to this specific area do not seem to be a priority. However, there is still some room for improvement given the high demanding nature that patients have towards these attributes, turning them into leverage opportunities that should be evaluated to optimize their benefits. Special attention should be given to *food variety* (g_{42}), the most influential subcriterion that is seen as first priority improvement. Despite displaying the highest satisfaction index (3.49%) out of the four subcriteria, the increased patient's demanding nature makes up a significant room for improvement that shall not be disregarded. *Volunteering staff*, with a criterion weight of 11.63%, is the second most influential patient satisfaction predictor. Patients are satisfied with the service provided by this personnel, but there is still an increased room for improvement, so these items are viewed as leverage opportunities that should be assessed with the respective priority. The Kano model classifies them as critical must-be requirements due to having a larger weight attributed by dissatisfied patients than by satisfied patients. *Attention of volunteering staff* (g_{92}), however, is classified as a high valued-added one-dimension requirement because satisfied patients allocate the same weight to this requirement as dissatisfied patients. This type of requirement leaves patients feeling satisfied when their expectations are met and dissatisfied if otherwise. *Obtained information* is seen as the third and last satisfaction predictor, having a weight of 11.44% (located above the criteria centroid). The results of MUSA revealed that *substitution in decision making*, and *anticipated vital will* (g_{14} and g_{15}, respectively) have weights equivalent to zero, and were thus, removed from the analysis. Regarding the remaining subcriteria, *patient's rights and duties*, g_{12}, has the highest overall weight (4.91%), but there is not a substantial difference between the weights of the three subcriteria. *Patient's guide* and *complaint means* (g_{11} and g_{13}, respectively) are critical must-be requirements, as a result of the weights associated with dissatisfied patients being higher than those associated with satisfied patients. Given that patients are not extremely demanding when it comes to these

Table 1. MUSA main results. W_{jk}: weight of criterion j and subcriterion j_k; S_{jk}: satisfaction index of criterion j and subcriterion j_k; D_{jk}: demanding index of criterion j and subcriterion j_k; Δ_{jk}: room for improvement of criterion j and subcriterion j_k.

Criteria	Subcriteria	W_{jk}	S_{jk}(%)	D_{jk} [-1;1]	Δ_{jk} [%]	Kano's model category
G_1	g_{11}	0.0340	3.1100	-0.0800	15.9675	Must-be, critical
	g_{12}	0.0491	4.6500	-0.0280	22.7219	Highly attractive
	g_{13}	0.0313	2.7900	-0.1200	14.7856	Must-be, critical
	global	0.1144	9.4400	0.0200	10.3601	Highly attractive
G_2	g_{21}	0.0058	0.5300	-0.0500	10.3051	Must-be, necessary
	g_{22}	0.0099	0.8100	0.3100	17.5963	Must-be, necessary
	g_{23}	0.0082	0.6500	0.1900	14.6144	Must-be, necessary
	g_{24}	0.0114	0.9000	0.1700	20.2858	Must-be, necessary
	g_{25}	0.0057	0.4000	0.2200	10.2688	Must-be, necessary
	g_{26}	0.0063	0.5300	0.0100	11.2799	Must-be, necessary
	g_{27}	0.0084	0.7000	0.0100	14.9546	Must-be, necessary
	global	0.0557	4.1600	0.2100	5.3383	Must-be, necessary
G_3	g_{31}	0.0297	2.1000	0.1900	32.2483	Must-be, critical
	g_{32}	0.0130	0.9400	0.1400	14.2250	Must-be, necessary
	g_{33}	0.0283	2.3300	-0.1200	30.6879	Highly attractive
	g_{34}	0.0191	1.6100	0.1000	20.9374	Less attractive
	global	0.0902	7.7900	-0.0600	8.3173	Less attractive
G_4	g_{41}	0.0316	1.9500	0.3500	24.5223	Must-be, critical
	g_{42}	0.0446	3.4900	0.0400	34.1066	Must-be, critical
	g_{43}	0.0259	1.7900	0.3800	20.1429	Must-be, critical
	g_{44}	0.0242	1.9500	0.2700	18.7766	Must-be, critical
	global	0.1262	9.2500	0.3400	11.4527	Must-be, critical
G_5	g_{51}	0.0072	0.5800	0.3300	8.9080	Must-be, necessary
	g_{52}	0.0087	0.6700	0.3900	10.7177	Must-be, necessary
	g_{53}	0.0083	0.6900	0.1800	10.1991	Must-be, necessary
	g_{54}	0.0149	1.0300	0.5300	18.2996	Must-be, necessary
	g_{55}	0.0128	0.9700	0.3900	15.7161	Must-be, necessary
	g_{56}	0.0201	1.4900	0.4400	24.6176	Must-be, critical
	g_{57}	0.0086	0.6300	0.4300	10.5630	Must-be, necessary
	global	0.0804	6.2600	0.4400	7.5367	Must-be, necessary
G_6	g_{61}	0.0100	0.8700	0.3900	11.1323	Less attractive
	g_{62}	0.0135	1.2200	0.2300	14.9158	Less attractive
	g_{63}	0.0136	1.2300	0.2000	15.0525	Less attractive
	g_{64}	0.0248	2.1400	0.1900	27.1855	Highly attractive
	g_{65}	0.0114	1.0500	-0.1100	12.6755	Less attractive
	g_{66}	0.0159	1.3000	0.2600	17.5982	Less attractive
	global	0.0892	7.9000	0.2500	8.2153	Less attractive
G_7	g_{71}	0.0204	1.8200	0.1800	23.9461	Highly attractive
	g_{72}	0.0221	1.9400	0.2200	25.8878	Highly attractive
	g_{73}	0.0154	1.4400	-0.0900	18.1646	Less attractive
	g_{74}	0.0257	2.4200	-0.2600	30.0351	Highly attractive
	global	0.0836	7.8100	-0.1500	7.7071	Less attractive
G_8	g_{81}	0.0257	2.0800	0.1700	28.8081	One dimensional, high valued added
	g_{82}	0.0330	2.7600	0.1400	36.7275	Highly attractive
	g_{83}	0.0130	1.0900	0.0700	14.7870	Less attractive
	g_{84}	0.0156	1.2600	0.2000	17.6251	One dimensional, low value added
	global	0.0873	7.5200	-0.1000	8.0735	Less attractive
G_9	g_{91}	0.0339	2.7300	0.2400	28.3348	Must-be, critical
	g_{92}	0.0402	3.4300	0.0100	33.3842	One dimensional, high value added
	g_{93}	0.0422	3.6100	0.1000	34.9896	Must-be, critical
	global	0.1163	9.0400	0.1100	10.5786	Must-be, critical
G_{10}	g_{101}	0.0105	0.8500	0.0800	13.9306	Less attractive
	g_{102}	0.0098	0.8700	-0.0800	13.0257	Less attractive
	g_{103}	0.0068	0.5600	0.2600	9.0689	Must-be, necessary
	g_{104}	0.0100	0.8000	0.2200	13.3226	One dimensional, low value added
	g_{105}	0.0154	1.3300	0.0500	20.2471	Less attractive
	g_{106}	0.0086	0.7900	-0.2400	11.3992	Less attractive
	g_{107}	0.0137	1.1500	0.1300	18.0401	Less attractive
	global	0.0748	6.1400	0.0800	0.0000	Less attractive
G_{11}	g_{111}	0.0428	8.8600	0.0100	47.5842	Highly attractive
	g_{112}	0.0391	3.0500	0.2800	46.3324	Must-be, critical
	global	0.0819	6.5700	0.3300	7.6519	Must-be, necessary
Subcriteria centroid		0.0192	1.6900	0.1475	20.2284	
Criteria centroid		0.0909	6.2900	0.1130	6.9621	

subcriteria and criterion, it can be of great interest to develop strategies that enhance the benefits of providing valuable information, with a special concern for *patient's rights and duties* that is somehow seen as surprise element that leaves patients feeling even more satisfied.

6.5. Managerial implications

When comparing the results from SEM and OLR, both methods consider *accommodations, auxiliary staff* and *exams and treatments* as influential. The only disparity is that SEM also deems *medical staff* as influential. However, when observing results from MUSA, discrepancies arise. *Food quality, volunteering staff and obtained information* are the predictors of patient satisfaction according to MUSA. Comparing these results with the results yielded from the literature review, it is possible to conclude that these outcomes are aligned with previous findings since all seven criteria deemed as influential are also designated as critical. As referred to earlier, the focus of healthcare is to provide the best care across all service dimensions. However, due to capital and resources restrictions, it is not always possible to implement prosperous changes, and compromises have to be made. Evaluating how patients perceive the service provided, and how satisfied they are about it is an effective approach that can help managers decide on how to allocate the available resources. In line with the results introduced above, there are seven dimensions (*accommodations, auxiliary staff, exams and treatments, medical staff, food quality, volunteering staff, and obtained information*) that deserve special attention when implementing health policies. It is suggested that health care managers implement periodic patient satisfaction surveys to supervise the impact that modifications might have on patient satisfaction. The focus of these modifications should be on factors such as the behaviour of health personnel (more specifically medical, auxiliary and volunteering staff), the registration of complaints, and the hotel characteristics of hospital management. It is important to remember that this is a case study in a specific internment setting, thus, the final results and eventual implications are only valid for this particular scenario.

7. Concluding remarks, limitations and directions for future research

The focus of this thesis was the evaluation of patient satisfaction through the implementation and comparison of different methodologies, which originated two scientific articles. One being a review article, and the other an empirical article with the methods' results. Factor analysis was conducted in a complementary nature to the remaining methods. From there, two separate analyses (depending on which dataset was going to be examined) emerged. SEM was applied to both analyses and conclusions were achieved considering common predictors. Nevertheless, bias is closely related to all aspects of SEM. For an estimator, a large amount of bias can turn a true positive value into a negative one, and vice versa. As explained, when performing SEM, the dataset has to follow a multivariate normal distribution to allow the usage of the ML estimator, otherwise, results will be biased. This assumption was verified for our dataset, yet a comparison of results using other estimation methods might be of interest. If an analysis is based on biased estimators, the results will obviously be misleading. This is a recommendation for future work, in order to eliminate bias that might be present in the results. Furthermore, some multivariate models are not suitable for dealing with ordinal scales, as is the case of SEM and factor analysis, methods that are still being used by some researchers (47% of the collected articles). They undertake mathematical operations that are not consistent with Stevens' theory and data categorization (Stevens, 1946), and should not be applied to this kind of data (Vieira et al., 2020). OLR was the following method to be implemented in the two analyses. Evidently, this method also has limitations that might add bias to results. OLR assumes that error variances are homoscedastic, meaning that they do not vary even if the value of the predictor changes. This assumption can hold or not, depending on the sample, but its inadequate usage might cause less precise results. Weighted regression, a method that allocates a weight (based on variance) to each data point might be the best way to replace heteroscedasticity with homoscedasticity (Frost, 2019). There is no indication that heteroscedasticity is present in the dataset, but its evaluation should be taken into account for future research. Lastly, MUSA was applied solely to the dataset of analysis B. The justification for this fact has already been presented, but it is based on the fact that MUSA can solely work with results from CPCA. This is a type of factor analysis that deals with categorical variables and can be applied to ordinal categorical data, such as the case of our dataset. Thus, for further research, it would be advantageous to implement CPCA, compare the outputs with the ones from PCA, and use the new dataset in a complementary nature with MUSA. The outputs of the MUSA method proved to be different from the other two methods, and such difference can rely on MUSA's assumption that criteria/subcriteria are independent of each other. In some cases, this assumption may not be met, given the presence of cognitive biasing effects, such as the Halo effect (a cognitive bias where people form an opinion about an attribute based on their impression on another attribute (Costa and Remedios, 2014). In order to correctly assess the interactions between criteria/subcriteria and diminish the Halo effect, MUSA-INT, that considers positive and negative synergies between attributes, could be applied (Angilella et al., 2014; Ferreira et al., 2018). This new method should be considered for future research. Another issue that might arise from the MUSA method is its sensitivity to the number of constraints and variables. Meaning that as a linear programming model, the higher the number of satisfaction levels (in this case, seven), the higher the probability of introducing instability and, ultimately, returning an infeasible model. However, as explained in past sections, a seven point-Likert scale seems to be ideal, so a compromise between the two conditions might be necessary. Once the methods' implementation was completed, a conclusion that *accommodations, auxiliary staff, exams and treatments, medical staff, food quality, volunteering staff, and obtained information* are satisfaction predictors was achieved. In the future, it would be interesting to target different patient groups, such as age, gender, comorbidities, and specifically for the internment service, medical speciality, and length of stay. For instance, patients with a longer stay base their judgement on different experiences that patients with a one-night stay. Such segmentation was not possible to execute with our dataset due to a reduced sample. However, since different patient groups value different service dimensions, data segmentation can provide more reliable results.

References

Abrantes, M. (2012). *Qualidade e satisfação: Opinião dos utilizadores de serviços de saúde hospitalares*. Universidade de Coimbra. https://estudogeral.sib.uc.pt/bitstream/10316/20506/2/tese_Qualidade%20e%20satisfa%c3%a7%c3%a3o%20formatada_completa%5b1%5d.pdf

Al-Abri, R. & Al-Balushi, A. (2014). Patient satisfaction survey as a tool towards quality improvement. *Oman Medical Journal*, *29*(1), 3–7.

Angilella, S., Corrente, S., Greco, S., & Słowiński, R. (2014). MUSA-INT: Multicriteria customer satisfaction analysis with interacting criteria. *Omega*, *42*(1), 189–200.

Byrne, B. (1994). *Structural equation modeling with EQS and EQS-Windows: Basic concepts, applications, and programming* (1st ed.). Sage Publications.

Cardozo, R. (1965). Experimental study and expectation. *Journal of Marketing Research*, *2*(3), 244–249.

Coltman, T., Devinney, T., Midgley, D., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262.

Diamantopoulos, A. (1999). Export performance measurement: reflective versus formative indicators. *International Marketing Review*, *16*(6), 444–457.

Donabedian A. (1988). The quality of care: how can it be assessed? *Journal of the American Medical Association*, *260(12)*, 1743–1748.

Evans, M., Jamal, A., & Foxall, G. (2009). *Consumer Behaviour* (2nd ed.). John Wiley & Sons, New York, USA.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Ferreira, D., Marques, R., Nunes, A., & Figueira, J. (2018). Patients' satisfaction: The medical appointments valence in Portuguese public hospitals. *Omega (United Kingdom)*, *80*, 58–76.

Ferreira, S. (2019). *A multiple criteria satisfaction analysis tool for measuring patients ' satisfaction with Portuguese secondary care : The case of internment services* (Issue October). Instituto Superior Técnico.

Grigoroudis, E. & Siskos, Y. (2002). Preference disaggregation for measuring and analyzing customer satisfaction: the MUSA method. *European Journal Operational Research*, *143*(1), 148–170.

Hekkert, K., Cihangir, S., Kleefstra, S., van den Berg, B., & Kool, R. (2009). Patient satisfaction revisited: A multilevel approach. *Social Science and Medicine*, *69*(1), 68–75.

Hilbe, J. (2017). *Logistic Regression Models* (1st ed.). Chapman and Hall, Tempe, USA.

Cox III, E. (1980). The optimal number of response alternatives for a scale: A Review. *Journal of Marketing Research*, *17*(4), 407.

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control*, *41*, 39–48.

Kaya, O., Teymourifar, A., & Ozturk, G. (2020). Analysis of different public policies through simulation to increase total social utility in a healthcare system. *Socio-Economic Planning Sciences*, *70* (December 2017), 100742.

Koletsi, D. & Pandis, N. (2018). Ordinal logistic regression. *American Journal of Orthodontics and Dentofacial Orthopedics*, *153*(1), 157–158.

Liu, X. & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data. *Journal of Modern Applied Statistical Methods*, *11*(1), 242–254.

Lovelock, C. & Wright, L. (2001). *Serviços: Marketing e gestão*. Saraiva, São Paulo, Brasil.

Mardia, K., Kent, T., & Bibby J. (1980). *Multivariate Analysis* (Z. W. Birnbaum & E. Lukacs (eds.); 1st ed.). Academic Press.

Marôco, J. (2014). *Análise de Equações Estruturais: Fundamentos Teóricos, Softwares e Aplicações* (2nd ed.). Lisbon, Portugal.

Muhtaseb, R., Lakiotaki, K., & Matsatsinis, N. (2012). Applying a multicriteria satisfaction analysis approach based on user preferences to rank usability attributes in E-tourism websites. *Journal of Theoretical and Applied Electronic Commerce Research*, *7*(3), 28–48.

Qin, H. & Prybutok, V. (2013). A quantitative model for patient behavioral decisions in the urgent care industry. *Socio-Economic Planning Sciences*, *47*(1), 50–64.

Rahmqvist, M. & Bara, A. (2010). Patient characteristics and quality dimensions related to patient satisfaction. *International Journal for Quality in Health Care*, *22*(2), 86–92.

Ramsaran-Fowdar, R. (2005). Identifying health care quality attributes. *Journal of Health and Human Services Administration*, *27*(4), 428–443.

Rietveld, T. & Van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Mouton de Gruyter.

Schoenfelder, T., Klewer, J., & Kugler, J. (2011). Determinants of patient satisfaction: A study among 39 hospitals in an in-patient setting in Germany. *International Journal for Quality in Health Care*, *23*(5), 503–509.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Thurstone, L. (1947). *Multiple factor analysis*. University of Chicago Press.

Valle, P., Mendes, J., Guerreiro, M., & Silva, J. (2011). Can welcoming residents increase tourist satisfaction? *Anatolia*, *22*(2), 260–277.

Vieira, I., Ferreira, D., & Pedro, M. (2020). Patient satisfaction with healthcare provision services: What does the literature tell us about it? Submitted to *Socio-Economic Planning Sciences*.

Vilares, M. & Coelho, P. (2011). *Satisfação e Lealdade do Cliente: Metodologias de Avaliação, Gestão e Análise* (2nd ed.). Escolar Editora, Lisbon, Portugal.

Vuković, M., Gvozdenović, B., Gajić, T., Stamatović Gajić, B., Jakovljević, M., & McCormick, B. (2012). Validation of a patient satisfaction questionnaire in primary health care. *Public Health*, *126*(8), 710–718.

Wang, T. & Ji, P. (2010). Understanding customer needs through quantitative analysis of Kano's model. *International Journal of Quality and Reliability Management*, *27*(2), 173–184.

Xiong, B., Skitmore, M., Xia, B., Masrom, M., Ye, K., & Bridge, A. (2014). Examining the influence of participant performance factors on contractor satisfaction: A structural equation model. *International Journal of Project Management*, *32*(3), 482–491.