A survey of evaluation methods used for holistic comfort assessment

Alessia Gadotti and Rossano Albatici

1 University of Trento, Department of Civil Environmental and Mechanical Engineering, Via Mesiano 77, 38123 Trento, Italy
alessia.gadotti@unitn.it

Abstract
The complexity of a subject such as indoor comfort has encouraged the development of evaluation methods and scoring systems aimed at presenting either multiple indicators in a reduced set of results, a score, or a single classification, which can be more comprehensible for all the stakeholders involved. This paper examines these existing indoor comfort assessment models, taking into consideration both objective and subjective methods. Their aggregating, weighting and rating systems are summarized and compared. Important issues have risen with such formalizations. A multilevel hierarchical structure seems to be most suitable for a structured evaluation but weightings need to be used to show the relative importance of the four commonly considered main criteria, namely thermal and visual comfort, acoustic and indoor air quality. There are numerous weights assignment techniques, some of the most common are the Analytical Hierarchy Process (AHP) and multivariate linear regression of occupant responses. However, weighting factors may vary between geographic regions, cultural conditions, and individual circumstances. In addition, there is the need to find a model that offers a formal and logical way to include qualitative values in the analysis. Furthermore, the interactions of factors or parameters at different levels are not all known and are not considered in the models here reviewed, although regarded as important. A survey of techniques for sorting and presenting performance values such as scorecard models, radar diagrams or overall indices, is also presented. The framework drawn from this review provides the basis for the formulation of a comprehensive formal evaluation model.

Keywords: Comfort, indicators, benchmark, assessment models, indoor environmental performance

1 Introduction
In order to evaluate indoor comfort in a building, research studies and practice procedures have mainly focused on the optimization of single components of the indoor environment (thermal, luminous and acoustic comfort and air quality), trying to identify objective relationships between parameters and resulting occupants reactions. This component-related approach focused on dose-response relations has led to well-known models, such as for thermal comfort (Fanger, 1982) and to regulations and guidelines (ISO, 2005; ANSI/ASHRAE, 2013). Only recently, an alternative concept has become established, approaching the indoor environment in a holistic way (Bluyssen, 2014). The European standard EN 15251 for the first time takes into consideration a range of different aspects, and others are also following this path (CEN, 2007; Nordic Standard, 2014). These regulations suggest an approach to classify and certify buildings using levels of individual environmental components, without however providing information about how to combine them into one index that can be used to classify indoor conditions. This scientific
Methodology toward the assessment of comfort has tried to combine the different components, into a compound measure of the built environment, its indoor environment and occupant responses as a complex system. In order to assess the occupants’ wellbeing, meaning health and comfort, several research methods have been developed. These models attempt to correlate the different indoor environmental factors with respect to the overall comfort or rather, as defined by Heinzerling et al., they take quality performance data and produce an evaluative numerical summary of the data (rating or score) (Heinzerling, et al., 2013). Two fundamental approaches can be distinguished – quantitative and qualitative – in which indoor environmental aspects are monitored directly via measurements or indirectly via questionnaires, respectively. The two typologies can be alternative or complementary: often both can be used in one study. In a review, Fassio et al. have classified two different main methods in assessing the indoor comfort: subjective-objective methods, which are a combination of qualitative and quantitative measurements, and objective methods, based only on quantitative measurements to be compared with a fixed set of criteria (Fassio, et al., 2014). However, a uniform measurement protocol has yet to be established and the individual models follow their own rules about space (sensor location) and time (measurement execution period). Organisations in the United States are attempting to standardise such a procedure with the development of a measurement protocol intended for commercial buildings (ASHRAE/USGBC/CIBSE, 2010), but an equivalent European document is still missing. Moreover, weighting schemes differ one to another, and it has yet to be established if and how it is possible to agree on a shared weights assignment. Furthermore, little agreement can be found between assessment scales and class thresholds, whose establishment is not properly justified. Additionally the impacts of the interactions between factors have not been considered yet by researchers and these are still addressed independently.

Research into ways of understanding how single factors contribute to, and interact with each other to result in a comprehensive comfort response from individuals is being actively developed, despite the complexity of the subject. Consensus is emerging that there is a need to improve assessment procedures and further research into them is required. To this end this paper aims to provide an understanding of how different existing comfort assessment models are constituted and a comparison of their aggregating, weighting and rating systems.

2 Comfort assessment models

A remarkable amount of variables and factors has an influence on the indoor environment. Commonly four main aspects of comfort are considered: thermal comfort, visual comfort, acoustic comfort and, of more recent interest, due to its health implications, indoor air quality. This is consistent with investigations showing that these four parameters are considered by occupants to be the most important in determining comfort (Frontczak, et al., 2012). Aesthetic quality and spatial and ergonomic quality, although significant, are not be considered in this review. This emerging definition of comfort can be seen as having a hierarchical structure composed by four levels (Figure 1). At the top level, the main strategy or fundamental goal is located, followed by four categories that are defined by a number of criteria, each one of them specified by one or more indicators. In this kind of formalization, it is usually recommended to go down until a measurable goal is reached. In this study, indicators are precisely defined as the measurable control parameters.
This formalization allows the complexity of the subject to be reduced, resulting in enhanced comprehensibility for different stakeholders. Comfort itself is divided into smaller parts which are individually analysed in order to better understand their contribution to the whole problem. Thomas L. Saaty claims that a hierarchical structure simulates the innate method of operation of the human mind. When facing a complex situation with a multitude of elements, the mind aggregates them into groups according to certain properties they share (Saaty, 1972). The main purpose of structuring the criteria is to form a basis for a formal evaluation process, taking into account all relevant objective and subjective factors.

In order to examine in depth different interpretations of this structure, a literature review was undertaken for papers presenting comfort assessment models or indoor environmental categories weighting scheme, using electronic database as Sciencedirect or Google Scholar, and a dozen research studies have been found and classified on the basis of aggregation methods.

2.1 Aggregation methods
In the examined subjective-objective methods, equations that attempt to predict occupant satisfaction based on objective measurements for each comfort category are provided. Satisfaction results represent the score (“sub-index”) for each individual comfort category. A linear relationship between perceived comfort and contributing categories is established and the overall index is expressed as a multivariate model, function of the sub-indices of each category multiplied by the weighting coefficients derived from regression results of questionnaires (Ncube & Riffat, 2012; Cao, et al., 2012). A similar procedure, but with a
non-linear relationship using multivariate logistic correlation, can be found in other studies (Lai, et al., 2009; Wong, et al., 2008; Mui, et al., 2015).

In the examined objective methods, a simple additive weighted method is used as aggregation method, where the parameters are measured on a common scale, multiplied by their respective weights, and added into an overall measure of quality (Chiang & Lai, 2002; Marino, et al., 2012; Reffat & Harkness, 2001). Objective measurements are compared with a fixed assessment class structure in which each range has a corresponding attribute score and the overall index is the combination of these sub-indexes through weighting process. Most of the green building certification tools use this kind of approach. Their protocols are structured in a similar manner as those in the cited comfort assessment models and are usually composed of the following elements: criteria or categories and indicators, which describe the performance of the defined criteria. They are also provided with an assessment scale that defines the requirements with which performance is designated as good or bad, using an allocation of scores and a weighting scheme to define the relative importance of criteria. In these models subjective measures can be taken but are not part of the assessment process. The data collected can be undertaken for validation and control purposes.

2.2 Weighting
To show the relative importance of the different criteria, weightings are used. Their assignment procedure is a crucial step in the multi-attribute analysis. Weightings can be elicited using hierarchical or non-hierarchical methods (Häkkinen, 2012). In the latter only bottom level weightings (indicator levels) are assigned, and upper levels (criteria) are derived as sums of weights of the indicators of which they are composed. More commonly, with a hierarchical method, weightings are established at all levels of the value tree and the weights at the bottom level are obtained by multiplying weights vertically. At all levels of the hierarchy the sum of the weights is expected to be equal to 1 (or 100%). A number of studies have attempted to prioritise indoor environmental categories. For the definition of the weighting factors, various possible methods have been applied. Some studies used regression coefficients of regression models obtained from subjective measurements to indicate the relative importance of the four categories to the overall comfort value (Lai, et al., 2009; Wong, et al., 2008; Ncube & Riffat, 2012). Other procedures harvest assumptions from expert forums or alternative structured methods. One example of formal weighting process is the Analytic Hierarchy Process in which the factors are judged by a panel of experts through pairwise comparisons: is thermal comfort strongly, slightly or equal important than indoor air quality? Paired comparison method and AHP are used in several studied models (Chiang & Lai, 2002; Liu, et al., 2012; Reffat & Harkness, 2001).

Among the analysed studies, thermal comfort has shown to be the most influential factor, slightly more important compared to indoor air quality and considerably more important compared to acoustic and visual comfort (Figure 2).

The advantage of AHP is first of all the possibility to include qualitative values in the analysis in a formal and logical way. The shortcoming of this well-known method is that AHP assumes that trade-offs between factors are linear functions and that there is no dependency among them, which is hardly true in the indoor environment. The criteria interact with each other rather than being independent. For example, interactions occur between thermal conditions and indoor air: odours are more annoying with higher temperatures and humidity than when the indoor climate is cool and dry.
In addition, the subjective judgment and preferences of decision makers have a great influence on the final adopted elicitation. To check the consistency of the judgments, an important part of the weighting assignment procedure is sensitivity analysis, which consists of studying how changes in weightings affect the results. In the case of assessment of building performance, this also helps to determine which are the most influential factors and how much the performance level of an indicator must be improved so that the building achieves a certain desired level of comfort. This element is present in most of the analysed models.

Figure 2. Summary of attempts at ranking indoor comfort categories (rank 1 = higher importance, rank 4 = minor importance).

2.3 Indicators
Quantitative indicators in general define the indoor performance of a building. Examples are temperature, ventilation rate, lighting intensity or pollutant concentration. The definition of these parameters occurs with the traditional “bottom-up approach” (Bluyssen, 2014), establishing threshold values after having defined dose-effect relationships. These dose-related parameters are frequently used in standards and regulations and in green building assessment tools, both national (such as ITACA, DGNB or HQE) and international (such as BREEAM, LEED, CASBEE or WELL). In Table 1, an analysis of the indicators used in different protocols is presented, considering several green building assessment tools and a number of the investigated models.
As shown, some protocols make use of a considerable range of indicators distributed over all the four main categories. Some indicators are very common among the examined assessment models, such as daylight factor, CO₂ concentration and background noise level. An attempt to review and select the most important indicator has been made within the European project Perfection (Performance Indicators for Health, Comfort and Safety of the Indoor Environment), focusing on the implementation in an indicator framework for the assessment of building performance (Huovila, et al., 2010).

### 2.4 Assessment scale and categories

Some studies refer directly to occupant satisfaction. Others, however, present a breakdown in assessment classes, categories or sanitary levels. Especially for objective models, comfort
ranges for each performance criterion are determined and a corresponding attribute value assigned (Figure 3). Indeed, a quantitative evaluation makes sense only if thresholds can be defined. Marino et al. present a breakdown in four quality classes for each category, using qualitative attribute and colours (I = green, II = yellow, III = orange, IV = red), while other studies assign to five categories a corresponding numerical value of 10, 8, 6, 4 and 2 (Chiang & Lai, 2002; Ncube & Riffat, 2012; Reffat & Harkness, 2001).

As seen for the previous weighting schemes, there is a high disagreement and variation between studies regarding these categories themselves and their limits and thresholds. Some indicators have ranges that vary widely between studies and categories not always reflect perceptible changes in occupants’ satisfaction (Heinzerling, et al., 2013).

### 2.5 Presentation of results

The results of a performance measurement may be presented in different ways. One of the most common representations consists of a matrix structure or scorecard to facilitate side-by-side comparison. The matrix could contain numerical score or qualitative attributes (e.g. high, medium and low quality) or the impacts are presented in its natural units, rather than being converted into a single measure of worth. In the latter case, the comparative rankings for a single impact can be indicated by colour coding. In this type of representation the table is showing both the rank-coding and numeric outcomes for the different impacts. The advantages of using also colour coding is the possibility of conveying an overall impression of how rankings change readily that can other ranking schemes. Scorecards can present some difficulties in identifying the general trends in performance.

Another appreciated graphical presentation of the results is a radar chart, also called star or spider diagram, which shows the individual categories scores providing an axis for each variable. The results can be presented in natural units showing the individual category scale on each axis. It is important to clarify that the scales are not directly comparable and comparison across variables is pointless in this case. When communicating results – this is...
also valid for other kinds of representation – it is advisable to provide both normalized indices and the original raw data. It is then recommended to normalize the results in order the criteria to be commensurate. An example of radar diagram with common scale between axes is the one presented in Figure 2.

An advantage of radar diagrams is that referenced performances (least acceptable, standard/mean or best practice) can be also indicated in the chart. However, despite this kind of representation that gives an immediate picture of the overall score through the polygonal shape area, no weighting of the criteria is included in it and comparison between buildings with similar overall performances but different individual category scores can be cumbersome.

The ultimate goal of many assessment models is to provide a single index. The advantages of presenting one score is that it can be tracked and analysed allowing building comparison and benchmarking, as easily as for energy consumption. The weakness of such approach is that inaccuracies in individual scores and assumption may be compounded during the process of combining multiple attribute ratings. Additionally, one can criticise the need to quantify quality, pointing out a contradiction in terms. However, a such a symbolic score is a good way to immediately scope out a situation (Peña & Parshall, 2012). Moreover, it helps to raise the profile of comfort in relation to the importance attributed to energy consumption in buildings. To this end, graphical representations inspired by typical energetic efficiency rating bars (a seven point scale, from A to G), as used by Marino et al., may contribute also to enhancing the perceived importance of the subject.

3 Discussion and emergent models

The reviewed related research indicates that the hierarchical structure is most suitably applied to the case of evaluating comfort since it enables multiple interdependencies to be tracked at different levels of detail, while at the same time providing a holistic overview of the system being analysed. Hierarchies or value trees are also flexible scheme for such studies, where additions and reductions may be easily implemented if they are well-structured. However, many questions are raised about these frameworks. Firstly, it is still not univocally established which weightings shall be best given to the different variables or if equal consideration should be given to each of them. Some research demonstrates that not every factor has the same importance, but questions remain on how to assign respective weights in a meaningful way. Some authors are sceptical about this framework and claim that it is impossible to attempt to introduce an index of overall satisfaction, based on different factors, since these factors would acquire different weightings under different circumstances (Humphreys, 2005; Bluyssen, 2014). The relative weightings of individual attributes change significantly over time, from culture to culture and from country to country. Both authors state that it seems prudent to continue consider each aspect separately instead of relying on an overall evaluation that cannot be stable or reliable. At this point in the discussion, it is important to point out that the purpose of the comfort assessment models being reviewed here is to provide a simplified classification method of the indoor comfort quality in order to enable comparison between individual buildings in terms of their overall environmental performance. This aim is argued here to be adequately served using a conventional hierarchy of comfort undertaken with reference to what the intended use of the environment might be.
Another issue requiring more work relates to the fact that no dependency between factors has been taken into account in these methods. Interactions inevitable occur not only between light and thermal comfort, or between thermal conditions and indoor air, but also even within one comfort category, for example, between different air compounds like VOCs or CO₂. Perception of comfort is also influenced by these interactions since our senses are more sensitive to certain indoor environmental stressors than others. How can this interconnectedness be practically or accurately included in a model? In order to answer to this question, further and more detailed research studies are needed, before being able to develop the required framework with indicators and interactions.

The third open-ended question comes from the observation that when such a method is applied in reality, reported satisfaction responses do not always work as expected. Satisfaction with the thermal environment within space depends as much on the space itself as on personal variables people bring to the area with them, such age, sex, gender and cultural conditioning. These parameters are still described with quantitative dose-related indicators, expressed in numbers, or ranges of numbers, assumed to be acceptable and healthy for people. However, these indicators are describing sensations of comfort, based on research directed merely at the detection of a stimulus in the environment. Perception of comfort, instead, refers to the way in which one interprets the information received from the environment. Subjective factors need to be taken into better account in further models, linking also psychological factors and influences.

Additionally, the assumption that the trade-offs between factors are linear functions is not fully accepted. Some researchers have suggested the use of Kano’s model to assess the level of human comfort and individuals’ satisfaction (Fekry, et al., 2014; Kim & de Dear, 2012; Martellotta, et al., 2016). Developed as model to evaluate consumer satisfaction in marketing and product development fields, Kano’s model starts from the assumption that the relationship between the performance of a product (or a space) and a user’s satisfaction with it is nonlinear and asymmetric (Figure 4). Using this model in the building context, indoor comfort factors can be classified into three categories: basic factors, bonus factors and proportional factors. The first ones can be seen as minimum requirements and can affect comfort perception in a negative way; the second ones are those beyond minimum expectation and can increase satisfaction; the latter ones change satisfaction levels proportionally to their own performance. Through multiple regression with dummy variables on subjective measurements, some basic and proportional comfort factors have been identified.

What emerges from the above-mentioned papers is that some factors have a “veto power” since when occupants show dissatisfaction towards one of them no combination of the other parameters can lead to satisfactory results in terms of overall comfort. This conclusion is particularly interesting as it stands opposite to theories based on a sort of trade-off between different indoor comfort aspects, or as called by Humphreys the forgiveness factor (Humphreys, 2005). In these authors’ opinion, the application of Kano’s model is of notable interest and further research needs to be done.
An alternative promising approach to describe correlations could be found in the implementation of artificial neural networks (ANN), a non-linear statistical data modelling tool. Inspired by the biological neural structure of the brain, this method can be used to model complex relations between inputs and outputs and can be suitable to simulate the changeable human decision-making process. So far the method has been applied as control strategy to HVAC system (Ari, et al., 2006) but a more useful (for this review purpose) application, with an ANN model fitted on objective measurements, have linked concentrations of pollutants in office with the observed number of workers presenting building-related symptoms (Sofuoglu, 2008). Durmisevic & Ciftcioglu (2010) also tried to predict the overall design performance of a large number of parameters in a healthcare environment. A similar approach could be implemented aimed at correlating the single attribute indoor comfort parameters to the overall comfort level perceived by the occupants with the use of a sufficient range of objective measurements.

4 Conclusions

When approaching a problem from a holistic point of view, the main risk is to get overwhelmed by an abundance of information. Comfort is a rather complex phenomenon, influenced by many factors involving both objective and also subjective issues, which may tempt many people to abandon the idea of measuring it. Moreover, the indoor environmental quality of a space, as perceived by its occupants, is often reported as not being acceptable even if standard quantitative requirements are met. However, it is possible to measure those factors that impact more significantly on comfort and to obtain a reasonable index, which, even if not perfect, is certainly much better than no comprehensive evaluation at all. The need to present an indicator for the overall indoor environment has been recognised also by a number of authors who suggest to take into
account all physical parameters, which should be weighted opportunely, based on how they are ranked in importance differently by different building users. Furthermore, the development of a holistic comfort classification system for use in certain building assessments could be required to be produced and presented along with the energy certification, to give comfort the visibility and significance it should be accorded when compared to energy consumption performance in summary reports used in building evaluations.

Many attempts have been studied in this review, showing that there is a need to align benchmarking, weighting and aggregation methods with the definition of an assessment protocol for the selected indicators, making use of a range and combination of measurement techniques to successfully evaluate the buildings indoor environments, based not only on external stressors but also on personal factors. The basis for the formulation of a comprehensive formal evaluation model should start from what emerged from these previous studies. By developing such a holistic model, it could be finally possible to present indoor environments in a more realistic way and to understand why certain environments perform better than others. In addition, such a tool could be used to support decision makers/designers in improving the performance of buildings in this respect or in assessing the outcome based on inputs for new design planning. In conclusion, a new approach is required, based on users’ satisfaction being measured comprehensively instead in terms of single components only, bringing back the focus in good design onto the primary goal of architecture, which is to create comfortable environments for building occupants.

References


**Copyright Notice**

Authors who submit to this conference agree to the following terms: Authors retain copyright over their work, while allowing the conference to place this unpublished work on the NCEUB network website. This will allow others to freely access the papers, use and share with an acknowledgement of the work's authorship and its initial presentation at this conference.