Energy & Buildings 337 (2025) 115676

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Design and validation of thermal comfort questionnaire using exploratory and confirmatory factor analyses

A. Luparelli [a], P. Papadopoulos [b], I. Kyprianou [b], S. Erba [c], A. Ingrosso [a], S. Carlucci [d,*]

[a] CETMA – European Research Center for Technologies Design and Materials, Brindisi, Cittadella della Ricerca, Italy
[b] Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus
[c] Department of Architecture and Urban Studies, Politecnico di Milano, Milan, Italy
[d] Department of Theoretical and Applied Sciences, University of Insubria, Varese, Italy

ARTICLE INFO

Keywords:
Thermal comfort
Survey design
Exploratory factor analysis
Confirmatory factor analysis

ABSTRACT

Within the field of indoor environmental quality, several objective and subjective techniques have been developed to monitor conditions and provide user feedback. One such approach features occupant-centred questionnaires, indicating different aspects of comfort, including thermal, visual, acoustic or related to air quality. This study introduces the development and validation of such a questionnaire to optimise its completion rate and meaningfulness. Exploration and confirmatory analysis techniques were applied to identify the questionnaire's underlying structures, and the models' reliability and validation secured the validity of our findings. Questionnaire factors appeared to be connected to different aspects of temperature and humidity, each accounting for a significant portion of response variability, indicating that the questionnaire items effectively capture multiple dimensions within these constructs. The proposed thermal comfort questionnaire was also found to be reliable, with room for improvement through additional testing of individual elements. Overall, the questionnaire serves its intended purpose and can be used in subsequent analysis to improve the indoor environments of building occupants. The validation logic presented here, an essential step in the process of monitoring and building improvements, can serve as a set of guidelines and be replicated by other studies employing subjective feedback collection in the indoor environmental quality realm.

1. Introduction

Gathering feedback from building occupants is crucial for assessing and improving indoor environmental quality. However, designing reliable questionnaires for collecting feedback on thermal comfort can be complex due to the subjective nature of comfort perception and the influence of various environmental and individual factors. Indeed, personal factors, cultural influences, and individual preferences significantly impact how individuals perceive and then, respond to the thermal environment [1]. While cross-cultural survey administration is crucial for ensuring the reliability and validity of thermal comfort questionnaires in diverse settings [2], there can be problems associated with administering survey instruments within specific populations (see, e.g., [3]). Conversely, low and unrepresentative response rates can affect the accuracy of survey results and introduce bias [4]. Various strategies can be employed to optimise response rates, such as using incentives and multi-mode survey techniques [4].

Another primary challenge in survey design is ensuring that the survey accurately captures the experiences and perceptions of the target population because using non-probability sampling methods can introduce bias and affect the representativeness of the survey results [4]. That requires careful consideration of the use of results and the characteristics of the target population [5]. Thus, crafting questions that capture this variability while maintaining objectivity poses a considerable challenge. For instance, the usage of Likert scales or semantic differential scales to gauge thermal sensation and preference has been widely employed, and they were included in the international standards ISO 10551 [6] and ISO 28802 [7]. However, the interpretation of these scales can be subjective, leading to varied responses and potential ambiguity in data analysis.

Moreover, the temporal dynamics of comfort perception add another layer of complexity. Occupants' comfort is not static but evolves over time due to seasonal variations, diurnal rhythms, and adaptive behaviours [8]. This temporal aspect demands survey instruments capable of capturing these fluctuations while ensuring minimal respondent burden

* Corresponding author.
E-mail address: salvatore.carlucci@uninsubria.it (S. Carlucci).

https://doi.org/10.1016/j.enbuild.2025.115676
Received 4 October 2024; Received in revised form 23 January 2025; Accepted 26 March 2025
Available online 2 April 2025

and fatigue. Longitudinal studies or periodic surveys can offer insights into seasonal variations and adaptive behaviours [9,10] if they can address the challenge of keeping target populations engaged over a long period [11,12].

In this regard, digital surveys of thermal experience can be used for comfort management. Clear et al. [13] shed light on the potential of digital surveys for capturing real-time thermal experiences and preferences. Understanding the effectiveness and limitations of digital survey tools is also essential for developing reliable questionnaires in the context of thermal comfort assessment.

Building on this theme, the scope of this study is to introduce a feedback questionnaire design to assess building occupant comfort and employ widely used statistical approaches to improve its reliability. To this end, exploratory and confirmatory factor analyses are utilised to uncover underlying structures of a large sample of empirical data related to personal acceptability, preference, evaluation and perception of the building occupants. The study is a valuable addition to the literature, providing guidelines for designing and verifying the validity of a thermal comfort questionnaire. It can be used by other researchers who wish to follow the same approach following consolidated statistical methods to assess and validate similar surveys or expand outside the thermal comfort field.

2. Best practices in questionnaire design and assessment

To properly formulate a statistical problem, scale development and statistical significance must be considered with care. In psychometrics, a measurement scale consists of a series of items used to detect a latent variable (e.g., user satisfaction and well-being, thermal comfort etc.) [14]. In principle, a latent variable is defined as such because it cannot be measured directly [15]; a measurement scale can assess its intensity through some observable variables, net of an error [16].

The measurement scale constitutes the tool the researcher uses to learn about the construct of interest. As such, it is essential to ensure that it is psychometrically robust, thus ensuring that the scale is a reliable and valid tool for research and analysis purposes [17,18].

2.1. How to design a measurement scale

In this section, we present the methodology used to develop the measurement scale of the post-occupancy evaluation (POE) questionnaire. To design scales that accurately measure latent factors through manifest items, it is crucial to adopt a rigorous methodological process. As the main tool capable to analyse the construct of interest, it is important to ensure that the measurement scale is adequately constructed in measuring what it is designed for and in providing statistically significant answers to the relevant research questions. The adequate measurement of abstract constructs – in fact, accuracy in the measurement of the constructs under consideration – is, indeed, one of the biggest challenges in questionnaire-based surveys [19].

Through the development of accessible and transparent scales, we aim to facilitate significant progress in the understanding of complex phenomena, such as subjective thermal comfort. According to DeVellis [17], eight steps can be followed to develop a measurement scale. The procedures mainly concern the definition of the construct of interest, domain definition, reliability, validity, and dimensionality of the measurement scale. The first five steps aim to construct the questionnaire, and the remaining three steps serve to finalise the measurement scale, ensuring that it possesses the property of reliability (Table 1).

Table 1
Guidelines in Scale Development [17].

Step	Description	Method and technique
Definition of construct	Clarify the construct to be measured	Use a well-defined theoretical framework to ensure that scale is appropriate to measure the desired construct, avoiding ambiguities
Generating item set	Create a diverse set of items that capture various aspects of the construct	It consists of writing the items and elaborating on the main features. Ensure that the items reflect the purpose for which the scale is designed, considering the appropriate number of items to cover the entire theoretical construct, taking care not to introduce unnecessary redundancies
Determining structure	Choose the structure for the measurement scale (e.g., Likert scales)	Choose between various scaling methods like Likert scales, semantic differential scales, etc. Consider the number of response categories
Expert review	Have the item set reviewed by experts to refine items	Feedback on the item set is obtained from a group of experts, who can help confirm the construct definition, refining or eliminating items
Validation items	Introduce validation items to determine the validity of the scale	Validation items are added to assess the validity of the scale to detect weaknesses (e.g. social desirability bias) and to assess construct validity through relationships with other constructs
Administration to subjects	Administer the constructed questionnaire to a preferably large sample	Aim for a sufficiently large sample size to ensure reliable analysis (∼300 participants)
Assessment of item reliability	Assess the reliability through inter-correlation and Cronbach's Alpha	Examine item-scale correlations, variances, and means. Utilise factor analysis to assess dimensionality and confirm internal consistency with reliability coefficients, such as Cronbach's Alpha
Optimising scale length	Balance between brevity and reliability to optimise the scale's length	Consider the effects of scale length on reliability. Use techniques like item analysis and calculating Cronbach's Alpha excluding each item to refine the scale

2.2. How to assess a measurement scale

When assessing the reliability and validity of a scale, it is crucial to ensure that generated data aligns with the research hypothesis and accurately represents the latent constructs under investigation. Factor analysis is an integral part of scale development and a rigorous statistical tool. It is used to confirm whether the set of test items can accurately and consistently identify underlying factors that explain the correlation patterns between observed variables, determining whether a questionnaire captures the concepts it is intended to measure [17,20].

Factor analysis is divided into two main techniques: exploratory and confirmatory. Exploratory factor analysis (EFA) is mainly used to uncover the underlying factor structure by associating one or more latent factors with a group of observed items [21]. We can then use EFA to explore how many dimensions of the latent construct could be reflected in the survey data and to examine further their reliability [17].

Confirmatory factor analysis (CFA) is a quantitative data analysis method that belongs to the family of structural equation modelling (SEM) techniques [22]. CFA is utilised to confirm whether the hypothesised factor structure is present in the survey data and to validate the hypothesised relationships between the observed variables and their underlying latent constructs from a background theory [17,22]. These analytical techniques are essential to confirm the reliability of the measurement scale, and the validity of the questionnaire used to measure the constructs of interest.

Therefore, the emphasis on these specific techniques in the assessment process is motivated by their ability to establish accuracy, consistency, and the theoretical alignment of the measurement scale with the construct it intends to measure. Moreover, an integral part of this process involves refining the scale by identifying and removing items that do not contribute effectively to the construct measurement so that the scale's broadest reproducibility can be ensured in different contexts [17].

2.2.1. Reliability of scale

Reliability is defined as the degree to which a measurement is free of errors (e.g., responses influenced by the respondent's mood or level of tiredness, etc.) and, thus, able to produce statistically significant results. A measurement scale can be said to be perfectly reliable when causal error is absent, i.e. when reflecting the underlying construct without being distorted by external or internal factors unrelated to the construct itself [23]. There are three popular forms for measuring scale reliability [17,18,24]: equivalence, internal consistency and stability (Table 2).

2.2.2. Validity of scale

Validity is a fundamental characteristic of a measurement scale, indicating how accurately the scale succeeds in measuring the latent construct it intends to quantify [17]. In other words, a scale is valid if the item it measures coincides with the phenomenon of interest, that is, if the scale measures what is intended to be measured. Table 3 presents the main techniques for measuring the validity of a scale [17,24,25].

3. Methodology

3.1. Questionnaire design

Thermal comfort is subjective to humans; therefore, Post-Occupancy Evaluation (POE) is used to assess the interactions between objective data (spot-measured, monitored) and subjective data (questionnaires, interviews, surveys). In this study, 1860 responses were collected and analysed (including missing or NaN values). Data collection took place in various buildings dedicated to different domains, including educational activities (1178), work (583), research (17), meetings or extracurricular activities (17), recreation and sports (12), health care (8) and activities related to daily life in living spaces (45). The participants in the study were mainly young people (children and adolescents, 1055), adults (756) and the elderly (32).

These data types are used to draw a comprehensive picture of the actual use and usability of the buildings, perceived as complex and interactive systems, including the physical building attributes, incorporated mechanical and other systems and equipment, and the actual use by their occupants. The POE questionnaires have been developed according to the standard ISO 10551 [6], which describes the subjective judgment scales, as reported in Table 4. This standard defines numerical grades for each subjective judgment scale. However, these numerical grades represent ordinal and not numerical variables. It is, therefore, important to avoid the collected data with these ordinal scales being mistakenly treated as numbers, as shown by Favero et al. [26].

The thermal comfort questionnaire is built to collect questions about participants' demographic data, thermal comfort-related data, and sources of discomfort. The demographic data section asks for info on age, sex, perceived body dimension, where there is more than a person in the space, metabolic activity, clothing resistance, and perceived health state. The thermal comfort-related data section asks for information about the right-here and right-now feelings of the participants regarding their thermal preference (indicated as *Thermal preference*), their judgment of the thermal environment (*Thermal acceptability*), how they find the temperature in this precise moment (*Thermal evaluation*), their perception of the thermal environment (*Thermal perception*), their preference regarding the humidity level (*Humidity preference*), their judgment of the humidity level (*Humidity acceptability*) and their evaluation of humidity level (*Humidity evaluation*). Table 5 details the ranges for each scale used and lists the items evaluated and utilised later in statistical analysis.

Adopting the scales presented in Table 5 and complemented by a set of questions to characterise the occupant, the questionnaire helps to (i) understand potential differences in thermal comfort perception and/or use of systems, (ii) understand if personal factors may affect the thermal response of a person, and (iii) assess the thermal indoor environment. The questionnaires have been tailored according to the use of each space (office or similar, health care, residence, school, sports facility).

3.2. Experimental design

To investigate the mentioned research questions, the experiment is explicitly designed to identify the relevant study group and monitoring variable. In the following sections, the study group design and the monitoring design are described in detail.

Table 2
Measuring scale reliability.

Reliability	Methods	Description
Equivalence	Inter-Rater	Assesses the degree of agreement among independent observers rating, coding, or assessing the same measurement. Agreement is evaluated using statistical tests like the percentage of agreement, Cohen's Kappa coefficient, Pearson correlation coefficient, and the intraclass correlation coefficient.
Internal consistency	Cronbach's Alpha	Measures the correlation among all items on a scale to determine their collective reliability in measuring a single concept or domain with minimal error. High internal consistency suggests a strong relationship between the items and the underlying latent variable.
	Split Half	Assesses reliability by dividing the items into two parts and comparing their internal consistency. This method is often associated with the Spearman-Brown coefficient for adjustment.
Stability	Test-Retest	Evaluates the consistency of results when the same individuals are measured under identical conditions at different times. Stability is often assessed by the correlation coefficient measuring the degree of association between item pairs and between an item and the total scale.

Table 3
Measuring validity.

Validity	Methods	Description
Face validity	Expert Assessment	Considered a qualitative preliminary check to ensure the measurement tool appears valid to observers (experts or target subjects). It assesses how well the items linguistically and analytically represent what is supposed to be measured
Content validity	Expert Panels	Indicates the extent to which a set of items comprehensively covers the domain of interest. It's usually assessed by a group of experts reviewing the scale's adequacy and completeness
Construct validity	Convergent	The high observed correlation between the measure to be validated and other measures of constructs theoretically related to the investigated construct
	Discriminant	There is little or no correlation between the measure to be validated and other measures of other constructs not related to the measured construct
Criterion validity	Concurrent	It involves assessing the measure's correlation with related constructs. If a scale score agrees with other valid measures (criterion variable) of the same construct, this is referred to as concurrent validity
	Predictive	It involves assessing its ability to predict future or related outcomes (predictive). If a scale score is used to get an estimate of the score in a criterion which is delayed in time, this is called predictive validity

Table 4
Subjective judgment scales according to EN ISO 10551.

Type of judgment	Perception	Evaluation	Preference	Personal acceptability	Personal tolerance
Wording	"How do you feel (at this precise moment)?"	"Do you find it…..?"	"Please state how you would prefer to be now"	"How do you judge this environment (local climate) on a personal level?"	"Is it …?"
Number of degrees	7 or 9 or 11 degrees	4 or 5 degrees	7 (or 3) degrees	2 category statement form or 4 degrees	5 degrees
Wording of degrees, e.g., for assessing thermal environments	- Extremely hot - Very hot - Hot - Warm - Slightly warm - Neutral - Slightly cool - Cool - Cold - Very cold - Extremely cold	- Comfortable - Slightly uncomfortable - Uncomfortable - Very uncomfortable - Extremely uncomfortable	- Much warmer - Warmer - A little warmer - Neither warmer nor cooler - Slightly cooler - Cooler - Much cooler OR - Warmer - No change - Cooler	"On a personal level, this environment is for me: - Acceptable rather than unacceptable - Unacceptable rather than acceptable OR - Clearly acceptable - Just acceptable - Just unacceptable - Clearly unacceptable	- Bearable/tolerable - Slightly difficult to bear/tolerate - Fairly difficult to bear/tolerate - Very difficult to bear/tolerate - Unbearable/ intolerable

Table 5
Item scales.

Items	Levels
Thermal preference	Lower < Without Change < Higher
Thermal acceptability	Clearly Acceptable < Just Acceptable < Just Unacceptable < Clearly Unacceptable
Thermal evaluation	Comfortable < Slightly Uncomfortable < Uncomfortable < Very Uncomfortable < Extremely Uncomfortable
Thermal perception	Cold < Cool < Slightly Cool < Neutral < Slightly Warm < Warm < Hot
Humidity preference	Lower < Without Change < Higher
Humidity acceptability	Clearly Acceptable < Just Acceptable < Just Unacceptable < Clearly Unacceptable
Humidity evaluation	Comfortable < Slightly Uncomfortable < Uncomfortable < Very Uncomfortable < Extremely Uncomfortable

3.2.1. Study population groups

Given the diverse building types included in the project pilot case studies, the population of interest is composed of:

- Adults (females and males) working in office annotated as adults;
- Residents (female and male) in a health care centre annotated as elderly and fragile people;
- Non-athletes (females and males, adults) training in sports facilities annotated as non-minors;
- Pupils (females and malescarlucci) studying in primary school within the age range 6 to 11;
- Pupils (females and males) studying in high school within the age range 14 to 18;
- Tenants (females and males, adults and young) living in private apartments.

All participants are volunteers and have provided their consent to participate in the survey.

Table 6
Demographics of case study buildings.

Country	Building name	Building type	Area (m^2)	No. of monitored zones	Zone use	Zone occupancy info	Nominal no. zone occupants
Norway	Eidet Omsorgsenter	Health care center	7039	5	4 flats and 1 shared office	Patients and staff	7
Norway	Ellingsøy Idrettshall	Sports center	2610	2	1 main hall and 1 dance room	Athletes and Dancers	20
Norway	Flisnes Barneskole	Elementary school	4477	2	2 classrooms	Student and Teachers	40
Norway	Hatlane Omsorgsenter	Health care center	5980	5	4 flats and 1 shared office	Patients and staff	7
Norway	Moa Helsehus	Medical center	2700	6	2 Individual offices, 1 break room, 3 shared offices	Patients and staff	14
Norway	Spjelkavik Ungdomsskole	High school	9700	5	4 classrooms, 1 shared office	Students and teachers	125
Norway	Tennfjord Barneskole	Elementary school	2490	7	6 classrooms and 1 shared office	Students and teachers	105
France	Green'ER building (G2Elab)	Campus building	700	8	5 shared offices, 1 student hall, 1 classroom, 1 laboratory	Professors, Officers, Post-Docs, PhD students	45
Cyprus	Guy Ourisson Building (GOB)	Campus building	2020	9	7 individual offices, 2 shared offices	Professors, Officers, Post-Docs	11
Cyprus	Graduate School (GS)	Campus building	580	6	1 classroom, 1 individual office, 3 shared offices	Officers, Post-Docs, PhD Students	37
Cyprus	Novel Technologies Laboratory (NTL)	Campus building	2440	4	2 individual offices, 2 shared offices	Professors, Officers, Post-Docs, PhD Students	13
Italy	C2 Tower	Residential Apartments	1250	4	Living room, bedroom	Residents	12
Italy	C3 Tower	Residential Apartments	1250	4	Living room, bedroom	Residents	12
Italy	C4 Tower	Residential Apartments	1250	4	Living room, bedroom	Residents	12

3.2.2. Case study

The case study involves fourteen (14) buildings that are either public, commercial, or private and located in four (4) countries (Norway, France, Cyprus, and Italy), covering a variety of climatic regions across Europe, different building types and typologies. Table 6 presents the demographics of the case study buildings, including type and total area of the building, number and use of monitored zones, and information and nominal number of zone occupants participating in this study. The data set used for the analysis consisted of responses to POE questionnaires distributed in different living spaces, such as classrooms, single or shared offices, private flats, and workshops, from September 1st, 2022, to November 10th, 2023.

4. Results and discussion

A total of 1,860 observations were collected, and after filtering out the non-numerical values (NaN), 1,769 observations were classified as valid. Subsequently, the dataset was randomly divided into two subsets (EFA: n = 884; CFA: n = 885). The responses were treated as ordinal variables to recognise levels of agreement/intensity without assuming a uniform distance between the different levels of agreement or intensity [27,28].

Factor analysis was then conducted to identify underlying latent factors associated with subjective thermal comfort.

4.1. Exploratory factor analysis

The Exploratory Factor Analysis (EFA) focused on thermal comfort variables, specifically thermal acceptability, evaluations, perceptions, preferences, and humidity-related measures. The analysis aimed to identify underlying latent constructs related to thermal perception through the examination of correlations between observable variables. Variables such as body dimension, presence of multiple occupants, sex, age, metabolic activity, clothing resistance, openings, and health state were excluded at this stage. These variables could play a crucial role in later stages of the analysis to control, stratify, or interpret the results obtained from the latent factors.

4.1.1. Correlations between items

A polychoric correlation matrix was employed to handle ordinal data [29–31] revealing relationships that might be underestimated by Pearson correlations [30,32]. The correlation values range from −1 to 1, where −1 indicates a perfect negative correlation, 0 implies no correlation, and 1 represents a perfect positive correlation. As shown in Fig. 1, thermal acceptability and evaluation (0.89) as well as humidity acceptability and evaluation (0.90) are strongly correlated, indicating close alignment in how respondents judge these conditions. A strong negative correlation between thermal preference and thermal perception (−0.78) suggests that individuals preferring higher temperatures tend to perceive the environment as colder, and vice versa.

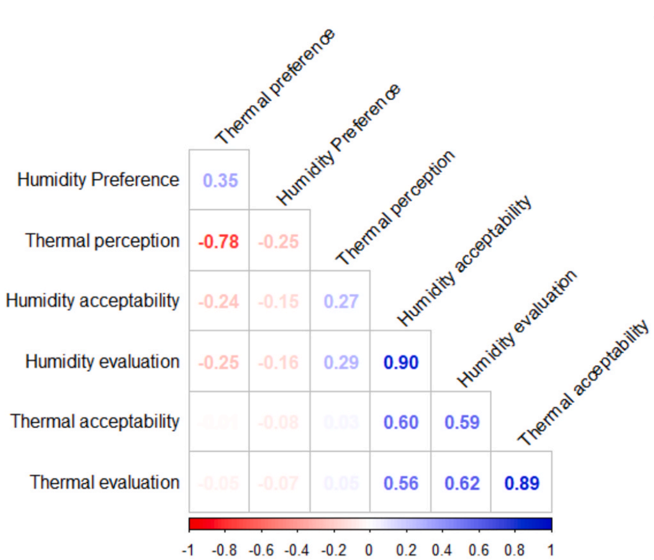

Fig. 1. Correlations between Items.

These strong correlations highlight a strong consistency between the way people evaluate and judge (and thus perceive) both the temperature and humidity of their living environment. That appears to be inversely correlated with how people prefer the actual temperature of their living environment.

4.1.2. Assessment of the suitability of the data

The first phase of factor analysis involves assessing the adequacy of the data through three tests: a) Kaiser-Meyer-Olkin (KMO), b) Bartlett's test and c) correlation matrix determinant [33–36]. These tests check, respectively: the common variance between items, the deviation from the identity matrix and the absence of multicollinearity, features necessary for the appropriateness of factor analysis.

4.1.2.1. Test for suitability of the data.

To detect the adequacy for factor analysis of the thermal comfort dataset, the KMO test is used. Similarly, Bartlett's test of sphericity, correlation matrix and determinant score are calculated. KMO (0.63) exceeded the 0.6 threshold but remains below the optimum value of 0.8 [33,34]. This indicates that the proportion of common variance among the variables is not extremely low and that the data could be adequate for factor analysis.

Bartlett's test was highly significant ($p < 0.001$), (Table 7), a p-value (5.990588E−88) was obtained, which means that observed correlation matrix is different from an identity matrix, and therefore, the data are suitable for factor analysis [34,35].

Finally, the value of the determinant of the correlation matrix obtained was 0.0068, which exceeds the threshold of 0.0001, indicating that there is no perfect multicollinearity between the variables. Since there is no perfect multicollinearity, factor analysis can be used [34].

Taking these results together, factor analysis seems feasible for the POE thermal comfort questionnaire data, albeit with some reservations due to the value just above the acceptable KMO factor adequacy analysis threshold.

4.1.3. How many factors to extract?

Another key step for factor analysis is to determine how many statistically significant factors to extract. A method used to determine the optimal number of factors to consider for performing a factor analysis is the scree plot obtained from parallel analysis [17,37]. Fig. 2 shows the eigenvalues plotted against the number of factors, representing the variance each factor explains. The parallel analysis results suggest extracting three statistically significant factors, as it corresponds to the highest number of factors where the eigenvalues of the actual data exceed the simulated ones.

4.1.4. Factor extraction

A factor analysis involves a two-stage process, initially the extraction of factors, followed by the rotation of these factors [38]. Factors obtained in the initial extraction phase are often difficult to interpret due to significant cross-loadings, where multiple variables load on multiple factors [34]. Consequently, factor rotation techniques are employed to

Table 7
Bartlett's Test of Sphericity.

Parameter	Value
Chi-Squared	477.92
p-value	0.00
Degrees of freedom	21

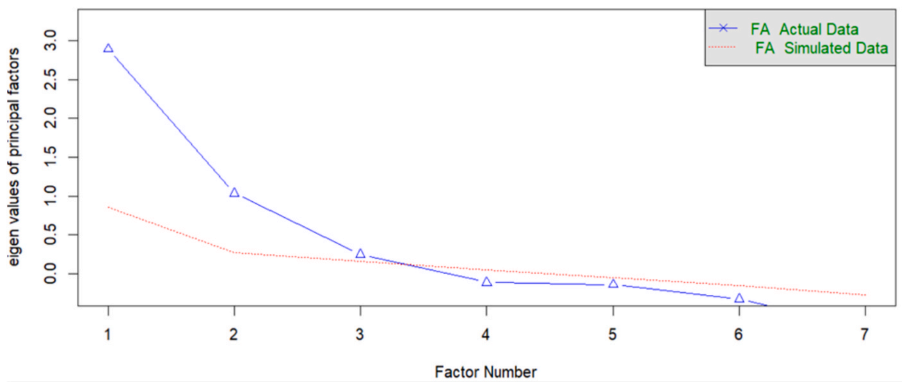

Fig. 2. Parallel Analysis Scree Plots.

improve factor interpretability. In this study, the Minimum Rank Factor Analysis (MRFA) was used to extract factors. MRFA was chosen for its ability to find communality estimates that best approximate the rank *r* solution, minimizing unexplained common variance and ensuring that all eigenvalues remain nonnegative [39,40]. That is particularly relevant when working with ordinal data, where polychoric correlations are employed [32].

Factor rotation techniques are either orthogonal (uncorrelated factors) or oblique (correlated factors). Varimax (orthogonal) was used as the rotation method, ensuring that each factor represents a specific and distinct sub-construct in the data [41].

4.1.5. Factor results and interpretation

The results of the factor analysis are shown in Fig. 3. The three latent factors (MRFA3, MRFA2, MRFA1) extracted via EFA are described below:

- MRFA3, Temperature-evaluation factor. It is characterised by high loads (0.9) for both *thermal acceptability* (i.e. acceptable or not) and *evaluation* (i.e. comfortable or not).
- MRFA2, Temperature-preference factor. It shows a strong positive load (1.0) for *thermal preference*, a negative load (−0.8) for *thermal perception*. The inverse relationship between these items indicates that individuals who prefer higher temperatures perceive the environment as colder, and vice versa. Item *humidity preference* has a positive loading to this factor, suggesting some degree of coherence between individuals' preference both of temperature and humidity, but the loading is weak and requires cautious interpretation.

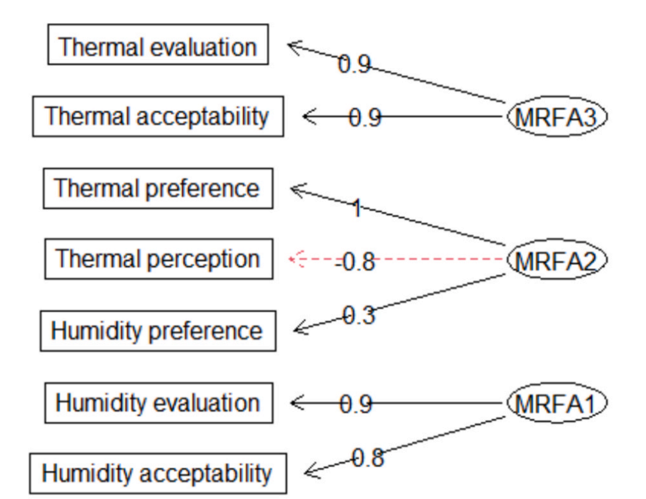

Fig. 3. EFA Factor loadings diagrams.

- MRFA1, Humidity-assessment factor. It has a high loading (0.9) for *humidity evaluation* and *acceptability* (0.8), indicating strong alignment between how people perceive a certain humidity level as comfortable (or not) also tend to judge the humidity as acceptable (or not).

4.1.6. Reliability of data gathered with the thermal comfort questionnaire

The internal consistency and stability of the items [42,44] of the questionnaire was assessed through the split-half method (Table 8). The split-half item is a method of assessing internal consistency in which the scale is divided into two parts, and the total score of each part is correlated with the other [43]. The results show a high overall reliability ($\lambda 4 = 0.93$, $\lambda 6 = 0.89$), an acceptable internal consistency with an α of Cronbach of 0.78, above the threshold of 0.7 [45] and appropriate inter-item correlations ($r = 0.34$, optimal range 0.20–0.40) [46].

The reliability could be improved by removing the item humidity preference (α would increase to 0.81). The low minimum split-half reliability ($\beta = 0.31$) suggests possible mismatches between some of the questionnaire items.

4.1.7. Variance explanation

To understand the importance and contribution of each factor in representing the information contained in the original data, the variance explained by the factors was analyzed (Table 9) [17].

SS loadings (MRFA3 = 1.89, MRFA2 = 1.80, MRFA1 = 1.73) indicate a balanced contribution of the three factors, with a slight predominance of MRFA3.

The three factors explain 27 %, 26 % and 25 % of the total variance, respectively, cumulatively reaching 77 %, which is generally considered satisfactory and indicates that the extracted factors can effectively capture the underlying structure of the data. The percentage of variance explained is, more or less, equally distributed among the factors (MRFA3: 35 %, MRFA2: 33 %, MRFA1: 32 %).

These findings suggest that all three factors (MRFA3, MRFA2, and MRFA1) appear to be connected to different aspects of temperature and humidity (perception, as well as evaluation, judgment and preference), each accounting for a significant portion of response variability.

Table 8
Split-half analysis.

Measure	Value
Maximum split-half reliability (λ4)	0.93
Guttman λ6	0.89
Average split-half reliability	0.77
Cronbach's α	0.78
Guttman λ2	0.82
Minimum split-half reliability (beta)	0.31
Average interitem *r*	0.34
Median interitem *r*	0.25

Table 9
Variance explained.

Measure	MRFA3	MRFA2	MRFA1
SS loadings	1.89	1.80	1.73
Proportion Var	0.27	0.26	0.25
Cumulative Var	0.27	0.53	0.77
Proportion Explained	0.35	0.33	0.32
Cumulative Proportion	0.35	0.68	1.00

4.2. *Confirmatory factor analysis*

A CFA was conducted to validate the three-factor structure identified by the EFA. Given the ordinal nature of the data, the Diagonally Weighted Least Squares (DWLS) estimator was used to estimate the relationships between the latent constructs and the associated observed items, for its accuracy in providing robust parameter estimates and more accurate standard errors for ordinal data or when the normality assumption is violated [47]. The model specifies three latent constructs:

- The *temperature assessment*. It includes the item's *thermal acceptability* and *evaluation*. This factor is related to how individuals evaluate (as comfortable or not) and judge (as acceptable or not) the temperature.
- The *temperature preference and perception*. It comprises the items' *thermal preference*, *thermal perception* and *humidity preference*. This factor can be interpreted as a combination of preference and subjective perception relating to temperature and humidity.
- The *humidity assessment*. It includes the item's *humidity acceptability* and *evaluation*. This factor represents the evaluative (as comfortable or not) and judgment (as acceptable or not) about humidity.

4.2.1. *Evaluation of the model*

The assessment of the fit of the CFA model was initially conducted by chi-square (χ^2) test. The statistical test ($\chi^2 = 25.475$, df = 11, $p = 0.008$) results in the rejection of the null hypothesis ($p < 0.05$) and suggests that the model does not fit the data perfectly [48]. However, given the known sensitivity of the χ^2 to sample size [49], additional fit indices were considered [50]. The goodness-of-fit indices of the CFA model are shown in Table 10.

The CFI and TLI values (both 0.999) significantly exceed the threshold of 0.95 [51,52], indicating an excellent model fit. The RMSEA (0.039, 90 % CI = 0.019–0.058) is below the threshold of 0.05 [51,52], confirming the goodness of fit. Similarly, the SRMR (0.027), which is particularly relevant for ordinal data, is well below the threshold value of 0.08 [53,54].

Overall, these indices confirm the adequate fit of the CFA model to the observed data.

4.2.1.1. Residuals correlation matrix. Residual correlation analysis was used to assess the adequacy of the CFA model in identifying relationships between variables [51,55]. Residuals represent the differences between observed correlations and those estimated by the model; values close to zero indicate adequate modelling of relationships, while values greater than 0.05 or 0.10 suggest relationships not adequately captured by the model [51,56].

The residuals of the correlation matrix (Table 11) are generally close to zero, confirming the goodness of fit of the model. The only exception is for the item humidity preference, which shows higher residuals: an underestimate in the relationship with thermal perception (0.07) and an overestimate in the relationships with humidity acceptability and evaluation (−0.06), suggesting difficulties in modelling the relationships of this specific item.

Table 10
The goodness of fit of the CFA model.

Metric	Value
Comparative Fit Index (CFI)	0.999
Tucker-Lewis Index (TLI)	0.999
Root Mean Square Error of Approximation (RMSEA)	0.039
Standardised Root Mean Square Residual (SRMR)	0.027

The distribution of the model residuals was evaluated by the Q-Q plot (Fig. 4), where only slight deviations from the theoretical normal line are observed, suggesting that the CFA model has a strong goodness of fit.

This result is confirmed by the Shapiro-Wilk normality test (Table 12), where a *W*-value of 0.96 and a *p*-value of 0.49 are returned, confirming that there is no statistically significant reason to reject the null hypothesis and believe that the model residuals deviate from normality.

4.2.2. *Convergent and discriminant analysis*

Construct validity was assessed by convergent and discriminant analysis [17]. Convergent analysis, based on standardized factor loadings [57], indicates a good representation of items in the respective factors when beta values exceed 0.70 [58]. The standardized factor loadings, reported in Table 13 and Fig. 5, show that:

- The Temperature-evaluation factor (T_{judge}) has strong associations for thermal acceptability ($\beta = 0.96$) and thermal evaluation ($\beta = 0.94$).
- The Temperature-perception factor (T_{perc}) has high loadings for thermal preference ($\beta = 0.98$) and thermal perception ($\beta = -0.73$), with the negative value indicating an inverse relationship between preference and perception. In contrast to the humidity preference item, which shows weak loading ($\beta = 0.38$).
- The Humidity-evaluation factor (H_{judge}) has strong convergence for acceptability ($\beta = 0.95$) and evaluation ($\beta = 0.96$).

Discriminant analysis was conducted using the Heterotrait-Monotrait Ratio (HTMT) [59,60], which compares correlations between different constructs (heterotrait) and within the same construct (monotrait). The accepted threshold value for discriminant validity is HTMT < 0.85 [59].

The results (Table 14) confirm a good differentiation between the constructs, specifically:

- Strong discriminant validity between thermal judgement (tmpj) and thermal perception (tmpp) (HTMT = 0.079)
- Good validity between thermal perception (tmpp) and humidity judgement (hmj) (HTMT = 0.350)
- Acceptable validity between thermal judgement (tmpj) and humidity judgement (hmj) (HTMT = 0.605).

The validity and reliability of the model were further assessed through the Average Variance Extracted (AVE) and the Composite Reliability (CR) (Table 15). The AVE, with an acceptability threshold of 0.5, assesses the variance explained by the latent factor with respect to measurement error, while the CR, similar to Cronbach's Alpha, indicates good reliability for values above 0.7 [58].

The thermal judgement (tmpj) and humidity judgement (hmj) factors show high AVE values (0.90 and 0.91) and adequate CR values (both 0.89), confirming strong convergent validity and internal consistency. The thermal perception factor (tmpp) shows an acceptable AVE (0.55) but an extremely low CR (0.04), suggesting adequate convergent validity but poor internal consistency.

Given the weak internal consistency observed for the humidity preference item, the CFA was conducted separately on two sub sample. Analysis confirming the factorial structure, although with differences in the way the humidity preference item is interpreted or evaluated by the participants.

In the Norwegian sample (Fig. 6), consisting mainly of children and young subjects, the item preference of humidity is associated with the

Table 11
Residuals correlations matrix.

Items	Thermal acceptability	Thermal evaluation	Thermal preference	Thermal perception	Humidity preference	Humidity acceptability	Humidity evaluation
Thermal acceptability	0,00	–	–	–	–	–	–
Thermal evaluation	0,00	0,00	–	–	–	–	–
Thermal Preference	0,02	−0,01	0,00	–	–	–	–
Thermal Perceptions	0,00	−0,01	−0,01	0,00	–	–	–
Humidity Preference	−0,03	−0,02	−0,01	0,07	0,00	–	–
Humidity Acceptability	0,03	−0,02	0,07	−0,02	−0,06	0,00	–
Humidity evaluation	−0,03	0,02	0,03	0,01	−0,06	0,00	0,00

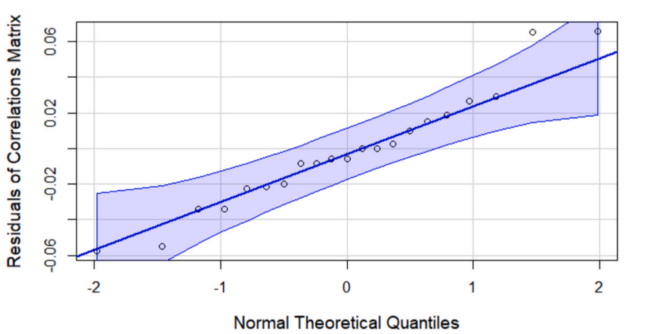

Fig. 4. Quantile distribution of the residuals.

Table 12
Shapiro-Wilk normality test.

Metric	Value
W-value	0.96
p-value	0.49

Table 13
Standardised Factor Loadings.

Latent Factor	Items	B	SE	Z	*p*-value	Beta
T_{judge}	Thermal acceptability	1.00	0.00	–	–	0.96
T_{judge}	Thermal evaluation	0.98	0.03	30.12	0.00	0.94
T_{perc}	Thermal preference	1.00	0.00	–	–	0.98
T_{perc}	Thermal perception	−0.75	0.05	−14.23	0.00	−0.73
T_{perc}	Humidity preference	0.39	0.04	10.98	0.00	0.38
H_{judge}	Humidity acceptability	1.00	0.00	–	–	0.95
H_{judge}	Humidity evaluation	1.02	0.03	37.25	0.00	0.96

perception of temperature, suggesting a difficulty in conceptualising humidity as a phenomenon distinct from temperature. The response pattern shows more simplified responses where temperature and humidity preferences move in the same direction, as well as greater randomness in the responses (evidenced by a low factor loading of 0.35 for humidity preference item).

In contrast, in the Cypriot sample (Fig. 7), composed mainly of adults, the humidity preference item shows a more logical relationship with the assessment of humidity. Although the factorial loading (−0.64) is below the conventional threshold, humidity preference shows an inverse relationship with humidity evaluation and acceptability, suggesting a greater conceptual consistency in the responses.

These results show a higher variability on humidity preference items between those two groups compared to the remainder of the items. This meaning that the humidity-related preferences might be influenced by different interpretative frameworks or subjective biases of the participants, reflecting in this case the different level of cognitive development between children and adults in understanding and evaluating thermal comfort parameters.

4.2.3. *Results summary*

Exploratory factor analysis (EFA) identified a well-defined trifactorial structure:

1. Thermal judgement factor (tmpj) is related to temperature assessment.
2. Thermal perception and preference factor (tmpp) is inversely related to the perception of room temperature.
3. Humidity judgement factor (hmj) is related to humidity assessment.

The factor structure was validated through confirmatory factor analysis (CFA), which showed a good fit to the data (CFI and TLI $>$ 0.95; RMSEA $<$ 0.05; SRMR $<$ 0.08).

Convergent validity analysis showed good convergence of the items with their respective factors (loading $>$ 0.7), with the exception of the humidity preference item. Separate analyses by sub-sample revealed that this criticality is particularly evident in the Norwegian sample (predominantly children), where the item is associated with the perception of temperature at low loadings, whereas in the Cypriot sample (predominantly adults) it shows a more consistent, although still below threshold (−0.64), relationship with the assessment of humidity. AVE values exceed the threshold of 0.5 for all factors, with particularly high values for tmpj and hmj, while tmpp shows a barely sufficient value.

The discriminant validity is confirmed by the moderate correlations between the factors ($<$0.85), although the correlation between tmpj and hmj deserves further investigation, considering the natural interaction.

5. Guidelines on thermal comfort questionnaire design and validation

This study presents a set of guidelines on the design and validation of a questionnaire investigating thermal comfort in indoor environments. Designing the survey with appropriate observable variables, constructs and measurement scales marks the first step; here the questions (in this case, on humidity and temperature) should reflect the scope of the study (thermal comfort assessment) while considering previous research and

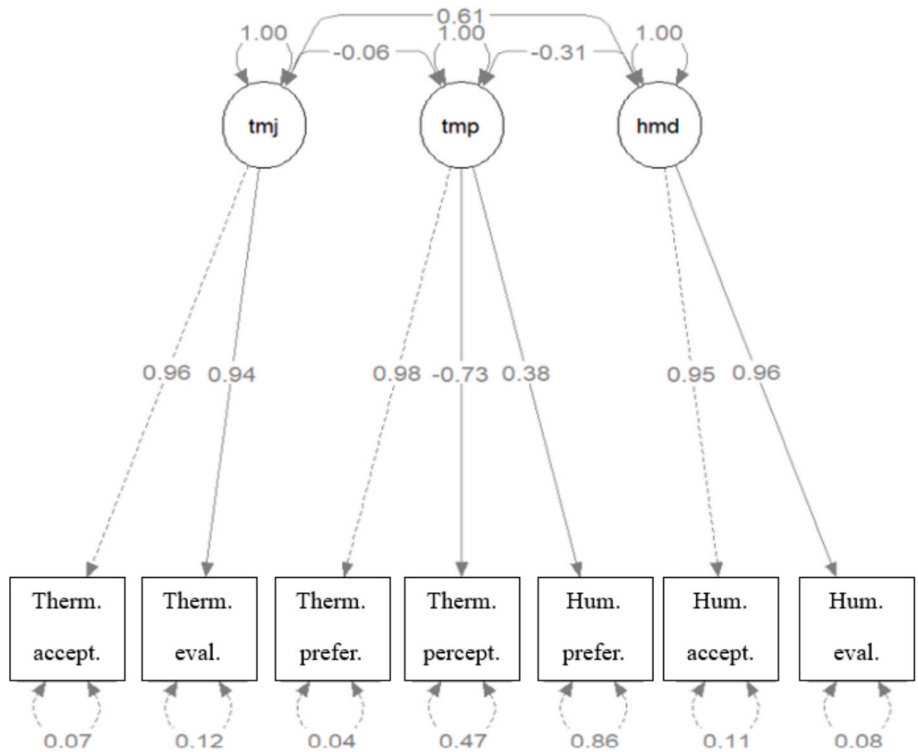

Fig. 5. CFA Diagram – Standardised Factor Loadings.

Table 14
Heterotrait–Monotrait ratio of correlations (HTMT).

Latent Factor	Thermal judgment (*tmpj*)	Thermal perception (*tmpp*)	Humidity judgment (*hmj*)
Thermal judgment (*tmpj*)	1.000	–	–
Thermal perception (*tmpp*)	0.079	1.000	–
Humidity judgment (*hmj*)	0.605	0.350	1.000

Table 15
Average Variance Extracted (AVE) and Composite Reliability (CR).

Index	Thermal judgment (*tmpj*)	Thermal perception (*tmpp*)	Humidity judgment (*hmj*)
AVE	0.90	0.55	0.91
CR	0.89	0.04	0.89

international standards. Administration of the questionnaires in the case study population groups is the next step, providing enough time to collect an adequate number of responses, which are filtered and checked for completeness.

Starting with exploratory factor analysis (EFA), we test whether the questionnaire measures latent constructs related to the observable variables defined earlier. The EFA is completed through robust statistical examinations of common variance, differences between the observed correlation matrix and an identity matrix and multicollinearity. When testing confirms the suitability of data, factors are extracted, and further testing of internal consistency and stability of items takes place. To complete the EFA, variance among the extracted factors should show that factors are connected to different aspects of the investigated domains (in this case, temperature and humidity), each accounting for a significant portion of response variability. This confirms that the questionnaire items effectively capture multiple dimensions within these constructs.

The next step is running the Confirmatory Factor Analysis (CFA), which can be seen as a hypothesis test in which the theoretical model is compared with the observed data. The CFA employs further consolidated statistical approaches, such as the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). When all indices conclude that the CFA model accurately represents the connections between the observed variables, residual correlation analysis determines the goodness of fit for the model. The last part of the CFA considers construct validity through convergent and discriminant analysis, which examine factor loadings and apply further indices to assess the extent to which items measure the same concept (convergent analysis), or the extent to which different latent factors represent distinct concepts (discriminant analysis).

Satisfactory statistical significance throughout this process determines the validity of the questionnaire, while less-than satisfactory significance highlights aspects that can be further improved. When time allows it, the observable variables can be re-designed to increase the statistical robustness of the survey, and special attention should be paid to acknowledging study limitations and mitigating them in subsequent research works.

6. Limitations

Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) validated the questionnaire's ability to measure different aspects of thermal comfort (judgements, evaluations, preferences and perceptions), confirming a consistent multifactorial structure. The results also revealed areas for improvement, particularly in the convergent and discriminant validity of some items, such as humidity preference.

However, the study has two main limitations. The first concerns the time dimension: despite the longitudinal nature of the data, the responses were treated as independent for pragmatic reasons related to the complexity of the data. A more in-depth analysis of factor variations

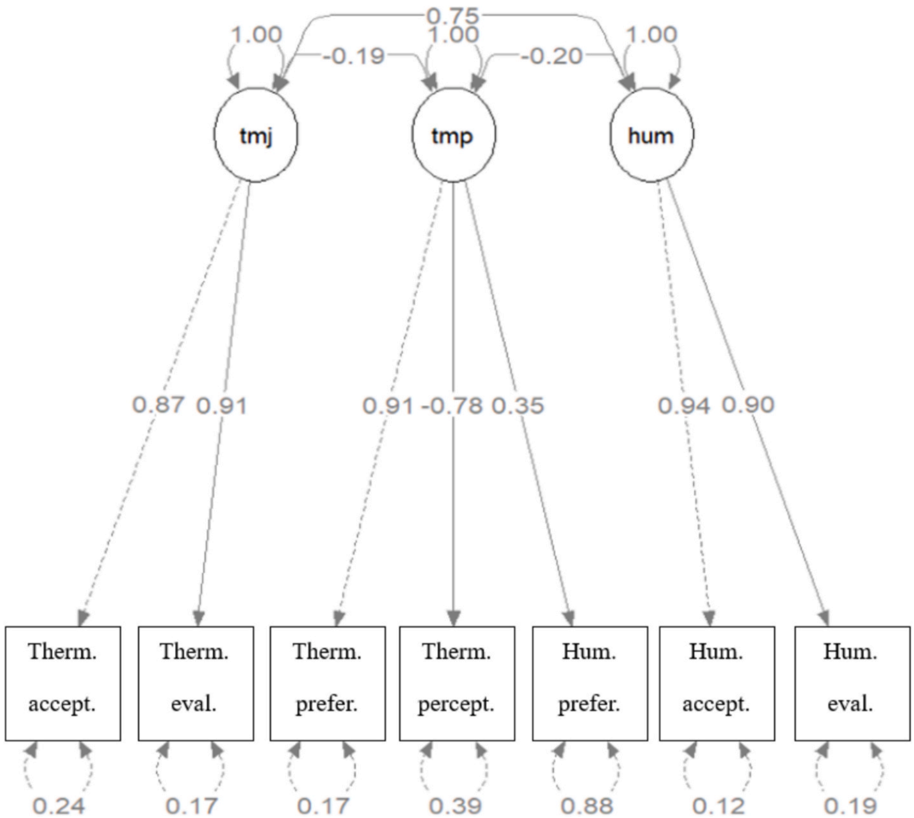

Fig. 6. CFA Diagram – Standardised Factor Loadings for Norwegian sample.

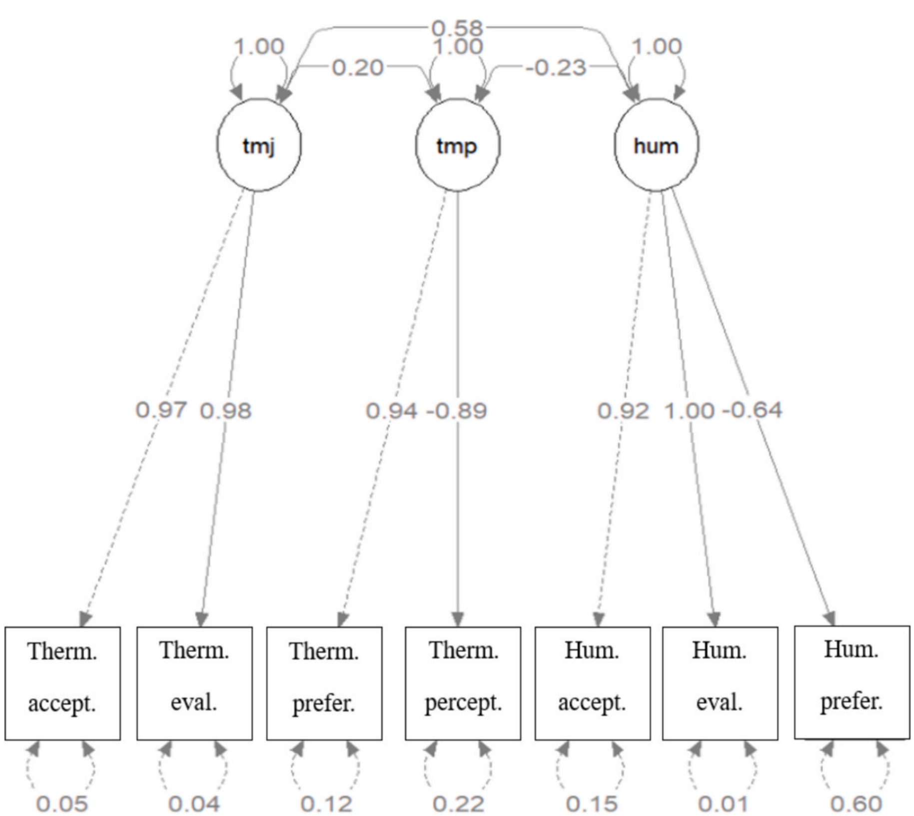

Fig. 7. CFA Diagram – Standardised Factor Loadings for Cypriot sample.

over time would require specific approaches for longitudinal data [61,62].

The second limitation concerns the invariance of the factor structure: while the analysis tested the overall validity of the questionnaire, it did not systematically examine how the structure of the latent constructs may vary as a function of key demographic characteristics such as age, gender, etc.

7. Conclusions

This study offers a few insights into the statistical approaches required to design an effective questionnaire on the topic of thermal comfort, which can allow the use of building occupants as consistent and reliable “sensors” of the thermal environment. Indeed, by means of exploratory and confirmatory factor analyses applied to a validated questionnaire, we have shown that the proposed survey represents the subjective opinions of building occupants in a scientifically sound way, showing its replicability. The Exploratory Factor Analysis showed that factors appeared to be connected to different aspects of temperature and humidity (judgment, evaluation, perception and preference), each accounting for a significant portion of response variability, indicating that the questionnaire items effectively capture multiple dimensions within these constructs. The Confirmatory Factor Analysis confirm and validates these latent factors, showing that some areas of this questionnaire need attention to improve convergent and discriminant validity.

The overall reliability of the thermal comfort questionnaire was adequate, with potential for improvement through in-depth evaluation of the individual items and implementation of the necessary adjustments where necessary. In particular, the humidity preference item showed critical convergent validity, with significantly different response patterns between the Norwegian (predominantly children) and Cypriot (predominantly adults) samples. This variability suggests that the interpretation of the humidity preference item may be influenced by specific subjective characteristics and support the importance of validating the questionnaire through pre-tests or pilot surveys, which allow the comprehensibility and effectiveness of the items and scales to be assessed in the target population.

In any case, the factorial model developed and specified seems to have a strong structure and good reliability in measuring the properties it is intended to measure. This work can serve as a set of guidelines in designing surveys to capture user feedback on the topic of thermal comfort, while future studies adopting similar methodological tools could benefit by using analytical approaches that can address longitudinal data, incorporating the temporal dimension in their analysis.

CRediT authorship contribution statement

A. Luparelli: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **P. Papadopoulos:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **I. Kyprianou:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Conceptualization. **S. Erba:** Investigation, Project administration, Validation, Writing – review & editing. **A. Ingrosso:** Software, Validation, Writing – review & editing, Investigation. **S. Carlucci:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization, Project administration, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the European Union’s Horizon 2020 research and innovation programme under a grant agreement for the COLLECTiEF (Collective Intelligence for Energy Flexibility) project (grant agreement ID: 101033683)

Data availability

The authors do not have permission to share data.

References

[1] M. Schweiker, S. Carlucci, R.K. Andersen, B. Dong, W. O’Brien, Occupancy and occupants’ actions, in: Explor. Occupant Behav. Build., Springer International Publishing, Cham, 2018, pp. 7–38, https://doi.org/10.1007/978-3-319-61464-9_2.
[2] M. Schweiker, et al., The Scales Project, a cross-national dataset on the interpretation of thermal perception scales, Sci. Data 6 (2019) 289, https://doi.org/10.1038/s41597-019-0272-6.
[3] J. Tindigarukayo, Challenges in conducting sample surveys in the Caribbean, Soc. Econ. Stud. 50 (2001) 167–188.
[4] J.A. Krosnick, S. Presser, K. Husbands Fealing, S. Ruggles, D.L. Vannette, The future of survey research: challenges and opportunities, 2015.
[5] OECD, Good Practices in Survey Design Step-by-Step, in: Meas. Regul. Perform., 2012. https://doi.org/10.1787/9789264167179-6-en.
[6] ISO, ISO 10551:2019 Ergonomics of the physical environment Subjective judgement scales for assessing physical environments, 2019.
[7] ISO, ISO 28802:2012 – Ergonomics of the physical environment – Assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses of people, 2012.
[8] J.F. Nicol, M.A. Humphreys, Adaptive thermal comfort and sustainable thermal standards for buildings, Energ. Buildings (2002), https://doi.org/10.1016/S0378-7788(02)00006-3.
[9] J. Kim, R. de Dear, T. Parkinson, F. Tartarini, P. Cooper, Longitudinal field study of thermal comfort in a low energy mixed-mode building, in: P. Rajagopalan, M.M. Andamon (Eds.), Engag. Archit. Sci. Meet. Challenges High. Density, 52nd International Conference of the Architectural Science Association, Australia, 2018.
[10] J. Malik, R. Bardhan, Thermal comfort perception in naturally ventilated affordable housing of India, Adv. Build. Energy Res. (2022), https://doi.org/10.1080/17512549.2021.1907224.
[11] K.J. McCartney, J. Fergus Nicol, Developing an adaptive control algorithm for Europe, Energ. Buildings 34 (2002) 623–635, https://doi.org/10.1016/S0378-7788(02)00013-0.
[12] H. Kim, H. Kang, H. Choi, D. Jung, T. Hong, Human-building interaction for indoor environmental control: evolution of technology and future prospects, Autom. Constr. 152 (2023) 104938, https://doi.org/10.1016/j.autcon.2023.104938.
[13] A.K. Clear, S.M. Finnigan, P. Olivier, R. Comber, ThermoKiosk: Investigating roles for digital surveys of thermal experience in workplace comfort management, in: Conf. Hum. Factors Comput. Syst. – Proc., 2018. https://doi.org/10.1145/3173574.3173956.
[14] D. Borsboom, D. Molenaar, Psychometrics, in: Int. Encycl. Soc. Behav. Sci., Elsevier, 2015, pp. 418–422, https://doi.org/10.1016/B978-0-08-097086-8.43079-5.
[15] K.A. Bollen, Latent variables in psychology and the social sciences, Annu. Rev. Psychol. 53 (2002) 605–634, https://doi.org/10.1146/annurev.psych.53.100901.135239.
[16] A. Borriello, S. Scagnolari, J.M. Rose, Reducing the randomness of latent variables using the evaluative space grid: Implementation in a hybrid choice model, Transp. Res. Part F Traffic Psychol. Behav. 62 (2019) 192–211, https://doi.org/10.1016/j.trf.2018.12.018.
[17] R.F. DeVellis, Scale Development Theory and Applications, fourth ed., SAGE Publ. 4 (2017) 256.
[18] G. Ursachi, I.A. Horodnic, A. Zait, How reliable are measurement scales? External factors with indirect influence on reliability estimators, Procedia Econ. Financ. 20 (2015) 679–686, https://doi.org/10.1016/S2212-5671(15)00123-9.
[19] F.F.R. Morgado, J.F.F. Meireles, C.M. Neves, A.C.S. Amaral, M.E.C. Ferreira, Scale development: ten main limitations and recommendations to improve future research practices, Psicol. Reflexão e Crítica 30 (2018) 3, https://doi.org/10.1186/s41155-016-0057-1.
[20] T.A. Kyriazos, A. Stalikas, Applied psychometrics: the steps of scale development and standardization process, Psychology 09 (2018) 2531–2560, https://doi.org/10.4236/psych.2018.911145.
[21] C.M. Woods, M.C. Edwards, 12 factor analysis and related methods, 2007, pp. 367–394. https://doi.org/10.1016/S0169-7161(07)27012-9.
[22] R.O. Mueller, G.R. Hancock, Factor analysis and latent structure analysis: confirmatory factor analysis, in: Int. Encycl. Soc. Behav. Sci., Elsevier, 2015, pp. 686–690.
[23] R. van Bork, M. Rhemtulla, K. Sijtsma, D. Borsboom, A causal theory of error scores, Psychol. Methods (2022), https://doi.org/10.1037/met0000521.
[24] AERA, Standards for educational and psychological testing, 1999.

[25] H. Taherdoost, Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research, SSRN Electron. J. 5 (2016) 28–36, https://doi.org/10.2139/ssrn.3205040.

[26] M. Favero, A. Luparelli, S. Carlucci, Analysis of subjective thermal comfort data: a statistical point of view, Energ. Buildings 281 (2023) 112755, https://doi.org/10.1016/j.enbuild.2022.112755.

[27] F. Franceschini, M. Galetto, M. Varetto, Qualitative ordinal scales: the concept of ordinal range, Qual. Eng. 16 (2004) 515–524, https://doi.org/10.1081/QEN-120038013.

[28] G.M. Sullivan, A.R. Artino, Analyzing and interpreting data from Likert-type scales, J. Grad. Med. Educ. 5 (2013) 541–542, https://doi.org/10.4300/JGME-5-4-18.

[29] U. Olsson, Maximum likelihood estimation of the polychoric correlation coefficient, Psychometrika 44 (1979) 443–460, https://doi.org/10.1007/BF02296207/METRICS.

[30] F.P. Holgado-Tello, S. Chacón-Moscoso, I. Barbero-García, E. Vila-Abad, Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables, Qual. Quant. 44 (2010) 153–166, https://doi.org/10.1007/S11135-008-9190-Y/METRICS.

[31] J. Moss, S. Grønneberg, Partial identification of latent correlations with ordinal data, Psychometrika 88 (2023) 241–252, https://doi.org/10.1007/S11336-022-09898-Y/FIGURES/4.

[32] J. Baglin, Improving your exploratory factor analysis for ordinal data: a demonstration using FACTOR, Pract. Assessment Res. Eval. 19 (2014) 1–15, https://doi.org/10.7275/dsep-4220.

[33] H.F. Kaiser, An index of factorial simplicity, Psychometrika 39 (1974) 31–36, https://doi.org/10.1007/BF02291575/METRICS.

[34] N. Shrestha, Factor analysis as a tool for survey analysis, Am. J. Appl. Math. Stat. 9 (2021) 4–11, https://doi.org/10.12691/ajams-9-1-2.

[35] M.S. Bartlett, Tests of significance in factor analysis, Br. J. Stat. Psychol. 3 (1950) 77–85, https://doi.org/10.1111/j.2044-8317.1950.tb00285.x.

[36] R.C. Rockwell, Assessment of multicollinearity, Sociol. Methods Res. 3 (1975) 308–320, https://doi.org/10.1177/004912417500300304.

[37] A.V. Crawford, S.B. Green, R. Levy, W.-J. Lo, L. Scott, D. Svetina, M.S. Thompson, Evaluation of parallel analysis methods for determining the number of factors, Educ. Psychol. Meas. 70 (2010) 885–901, https://doi.org/10.1177/0013164410379332.

[38] S.H. Teoh, A.C. Koo, P. Singh, Extracting factors for students' motivation in studying mathematics, Int. J. Math. Educ. Sci. Technol. 41 (2010) 711–724, https://doi.org/10.1080/00207391003675190.

[39] A. Shapiro, J.M.F. ten Berge, Statistical inference of minimum rank factor analysis, Psychometrika 67 (2002) 79–94, https://doi.org/10.1007/BF02294710.

[40] J.M.F. ten Berge, H.A.L. Kiers, A numerical approach to the approximate and the exact minimum rank of a covariance matrix, Psychometrika 56 (1991) 309–315, https://doi.org/10.1007/BF02294464.

[41] H.F. Kaiser, The varimax criterion for analytic rotation in factor analysis, Psychometrika 23 (1958) 187–200, https://doi.org/10.1007/BF02289233.

[42] H.G. Osburn, Coefficient alpha and related internal consistency reliability coefficients, Psychol. Methods 5 (2000) 343–355, https://doi.org/10.1037/1082-989X.5.3.343.

[43] R.L. Johnson, J. Penny, Split-Half Reliability, in: Encycl. Soc. Meas., Elsevier, 2005, pp. 649–654, https://doi.org/10.1016/B0-12-369398-5/00096-7.

[44] L.J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika 16 (1951) 297–334, https://doi.org/10.1007/BF02310555.

[45] K.S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. 48 (2018) 1273–1296, https://doi.org/10.1007/s11165-016-9602-2.

[46] R.L. Piedmont, Inter-item correlations, in: Encycl. Qual. Life Well-Being Res., Springer Netherlands, Dordrecht, 2014, pp. 3303–3304, https://doi.org/10.1007/978-94-007-0753-5_1493.

[47] C. DiStefano, H.L. McDaniel, L. Zhang, D. Shi, Z. Jiang, Fitting large factor analysis models with ordinal data, Educ. Psychol. Meas. 79 (2019) 417–436, https://doi.org/10.1177/0013164418818242.

[48] C.D. Nye, Reviewer resources: confirmatory factor analysis, Organ. Res. Methods 26 (2023) 608–628, https://doi.org/10.1177/10944281221120541.

[49] M. Alavi, D.C. Visentin, D.K. Thapa, G.E. Hunt, R. Watson, M. Cleary, Chi-square for model fit in confirmatory factor analysis, J. Adv. Nurs. 76 (2020) 2209–2211, https://doi.org/10.1111/JAN.14399.

[50] M.A. Babyak, S.B. Green, Confirmatory factor analysis: an introduction for psychosomatic medicine researchers, Psychosom. Med. 72 (2010) 587–597, https://doi.org/10.1097/PSY.0b013e3181de3f8a.

[51] N.K. Bowen, S. Guo, Evaluating and improving CFA and General structural models, in: Struct. Equ. Model., Oxford University Press, 2011, pp. 135–166. Doi:10.1093/acprof:oso/9780195367621.003.0006.

[52] Y. Xia, Y. Yang, RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods, Behav. Res. Methods 51 (2019) 409–428, https://doi.org/10.3758/s13428-018-1055-2.

[53] D. Shi, A. Maydeu-Olivares, Y. Rosseel, Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA, Struct. Equ. Model. 27 (1) (2020) 1–15, https://doi.org/10.1080/10705511.2019.1611434.

[54] L. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, Struct. Equ. Model. A Multidiscip. J. 6 (1999) 1–55, https://doi.org/10.1080/10705519909540118.

[55] A. Maydeu-Olivares, D. Shi, Effect sizes of model misfit in structural equation models, Methodology 13 (2017) 23–30, https://doi.org/10.1027/1614-2241/a000129.

[56] R.B. Kline, Principles and Practice of Structural Equation Modeling, fourth ed., The Guildford Press, 2016.

[57] S.C. Dunn, R.F. Seaker, M.A. Waller, Latent variables in business logistics research: scale development and validation, J. Bus. Logist. 15 (1994).

[58] G.W. Cheung, H.D. Cooper-Thomas, R.S. Lau, L.C. Wang, Reporting reliability, convergent and discriminant validity with structural equation modeling: a review and best-practice recommendations, Asia Pacific, J. Manag. (2023) 1–39, https://doi.org/10.1007/s10490-023-09871-y.

[59] J. Henseler, C.M. Ringle, M. Sarstedt, A new criterion for assessing discriminant validity in variance-based structural equation modeling, J. Acad. Mark. Sci. 43 (2015) 115–135, https://doi.org/10.1007/s11747-014-0403-8.

[60] J.F. Hair, G.T.M. Hult, C.M. Ringle, M. Sarstedt, N.P. Danks, S. Ray, Evaluation of reflective measurement models, Springer, Cham, 2021, pp. 75–90. doi:10.1007/978-3-030-80519-7_4.

[61] M.C. Corballis, R.E. Traub, Longitudinal factor analysis, Psychometrika 35 (1970) 79–98, https://doi.org/10.1007/BF02290595.

[62] U. Olsson, L.R. Bergman, A longitudinal factor model for studying change in ability structure, Multivariate Behav. Res. 12 (1977) 221–241, https://doi.org/10.1207/s15327906mbr1202_8.