WINDSOR 2020

On the linguistic challenges of cross-national research in thermal comfort: The effects of language choices in Greek and Swedish thermal perception questionnaires used in two large-scale surveys conducted two decades apart

Eleni Ampatzi[1], Despoina Teli[2] and Marcel Schweiker[3]

[1] Welsh School of Architecture, Cardiff University, Cardiff, UK, ampatzie@cf.ac.uk.
[2] Division of Building Services Engineering, Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden, teli@chalmers.se
[3] Building Science Group, Karlsruhe Institute of Technology, Karlsruhe, Germany, marcel.schweiker@kit.edu.

Abstract: Recent analyses on translations of the widely used ASHRAE scale have highlighted differences between the interpretation of thermal perception scales attributed to contextual factors. This study examines the influence of language choices for thermal perception scale anchors used in questionnaire studies. It focuses on the translations of the subjective scales for thermal perception in the Swedish and Greek languages, which differed between the two surveys. The first set of translations and datasets belongs to the Smart Controls And Thermal Comfort project whilst the second set belongs to the surveys conducted within the framework of the International Energy Agency - Energy in Buildings and Communities Program Annex 69 "Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings". This paper presents the findings of a comparative analysis between the two datasets in relation to the wording of the scales and the resulting category widths. The method of successive categories is used to estimate the psychological widths of the scale categories in order to examine the scales' behaviour. The paper discusses linguistic and climatic factors that may influence the way the translations are perceived and used.

Keywords: thermal sensation, scales translation, psychological continuum, successive categories

1. Introduction

The widely used thermal perception (sensation, comfort, preference, acceptance) scales have been developed in English and are included in international standards on the indoor environment, i.e. ASHRAE standard 55 (ASHRAE, 2017) and ISO 17772 (CEN, 2017). International standard ISO 10551 (CEN, 2019) provides guidance for the construction of appropriate subjective scales for the evaluation of the physical environment, based on the English version. The use of translated versions of such subjective scales, in surveys with native speakers, is meant to limit the risk of linguistic misunderstandings and improve the outcomes' reliability and relevance. It has however become evident that the translated versions of a scale do not always maintain the meaning and assumptions of the original (Rijal, 2012; Pitts, 2020; Schweiker *et al.*, 2020). For example, recent analyses on translations of the widely used ASHRAE scale have highlighted differences between the interpretation of thermal perception scales (Humphreys et al, 2016; Khatun et al., 2017; Schweiker et al., 2020), some of which are attributed to contextual factors. This creates problems in cross-national research that aims to unearth factors, other than language, that may influence subjective perceptions and votes.

Therefore, this paper discusses linguistic and climatic influences on the way the translations are perceived and used.

1.1. Aim

The aim of this work is to examine the behaviour of two sets of such scales developed and tested in two separate research projects. Out of the range of languages involved in both projects, this work focuses on the Swedish and Greek versions to examine in particular a) within each set whether the assumed equivalence of neutrality and preference for 'no change' is indeed present, and what this may indicate, and b) how the scales' category widths and their distribution differ. The parallel analyses of two different language sets provides the opportunity to make preliminary observations of -other than linguistic- factors that may have further influenced the scales' behaviour in the course of the last two decades.

2. Methods

2.1. The method of successive categories

In line with previous studies (Humphreys et al, 2016; Williamson et al, 2018; Al-Khatri et al, 2019), for each language version the method of successive categories is used to establish the psychological widths of the scale categories. The method of successive categories was originally developed by Louis Thurstone and Rensis Likert for the analysis of psychophysics (Guilford, 1954). It was introduced to thermal comfort research by Humphreys et al (2016) and has since been used for examining the changing thermal comfort expectations with time (Williamson et al, 2018) and the behaviours of translated thermal perception scales in Arabic (Al-Khatri et al, 2019). The application of the method in this area of research relies on the assumption that the responses on warmth and coolness in thermal sensation scales and the equivalent responses in thermal preference, acceptance and comfort scales exist on a 'psychological continuum', i.e. the series of responses form a straight line, "signifying changes in a single direction" (Guilford, 1954). The psychological continuum is assumed to represent the response to a physical (stimulus) continuum that describes the change of physical properties that cause that response. Despite the fact that the 'psychological' data is ordinal, there is the notion that it can approximate a continuous distribution which– when the data is of sufficient size- is expected to tend to the Normal form, in line with the Central Limit Theorem (Humphreys et al, 2016).

The method starts with the calculation of cumulative proportions for each scale category, followed by their transformation into Probits, which when plotted represent the upper category boundaries. The scale can then be renumbered by estimating each (apart from the end extremes) category's middle point, as the mean of its upper and lower margins. This process allows i) juxtaposing and examining the relations between different scales used within the same survey and ii) transforming such ordinal scales to interval scales for further statistical analyses. In this paper we apply the method to four datasets, and primarily examine [as per (i) above] the relationships between thermal sensation and preference scales in each case, aiming to establish the usefulness of the existent scales' translations in providing researchers with a true account of a population's thermal experience.

2.2. The surveys and scales' translations

This study uses two large-scale cross-national questionnaire studies conducted two decades apart. It specifically focuses on the translations of the subjective scales for thermal sensation and preference from English to Swedish and Greek, as these translations differed between the two surveys (see also Table 1).

The first set of translations and datasets belongs to the *Smart Controls And Thermal Comfort* (hereafter referred to as SCATs) project that ran in the period 1996 to 1998 and formed the basis of the adaptive comfort model in the European standard EN 15251. The SCATs project has provided the research community with a large dataset of multi-variate comfort information collected from native speakers in the UK, Sweden, Greece, France and Portugal (McCartney and Nicol, 2002; Ličina et al., 2018). The transverse questionnaire used (administered monthly) included thermal sensation (TS) and thermal preference (TP) scales for the evaluation of the parameter 'temperature'. The questionnaires were first produced in English and then translated into the four additional languages. To minimise the chance of inappropriate wording choices in the translations, a standard procedure was followed, with a subject-expert translating each scale into their native language and a non-expert, who was a native speaker of the particular language and fluent in English, independently retranslating them back to English (Humphreys et al, 2016). The translations were then further improved if discrepancies between the retranslation and the original English version were observed (Nicol at al, 2000). The project recruited in total around 850 participants.

The second set of translations and datasets belongs to the surveys conducted in 2017/18 within the framework of the International Energy Agency - Energy in Buildings and Communities Program (IEA-EBC) Annex 69 "Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings" (hereafter referred to as Annex 69) with the aim of assessing the perception of thermal comfort scales. This dataset and its analyses have recently been published (Schweiker *et al.*, 2019)(Schweiker *et al.*, 2020). For this study, 20 additional language versions of the 'same' questionnaire, originally conceived in English and using verbal anchors[1] from ISO 10551, were prepared and used. The questionnaire was translated in the other languages by participating subject-experts. Pilot tests were performed with at least seven individuals (laypersons and experts) to ensure that questions were perceived as intended and revisions were made following these tests if needed.

Table 1: Translations of verbal anchors for thermal sensation scale in SCAT and Annex 69.
** denotes language type, as defined in* (Schweiker *et al.*, 2020) *and reproduced here in Table 2.*

English ASHRAE SCAT & Annex 69	**Greek (SCAT)**	**Greek (Annex 69)**	**Swedish (SCAT)**	**Swedish (Annex 69)**
Cold (-3)	Ελαφρύ κρύο	Κρύο	Kall	Mycket kall
Cool (-2)	Πολύ δροσερά	Δροσερά	Sval	Kall
Slightly cool (-1)	Αναπαυτικά λίγο δροσερά	Λίγο δροσερά	Något sval	Något kall
Neutral (0)	Άνετα ιδανικά	Ουδέτερα	Neutral	Varken varm eller kall
Slightly warm (1)	Αναπαυτικά λίγο ζεστά	Λίγο ζεστά	Något varm	Något varm
Warm (2)	Πολύ ζεστά	Ζεστά	Varm	Varm
Hot (3)	Πάρα πολυ ζεστά	Πολύ ζεστά	Het	Mycket varm
2*	3c*	3c*	2*	1*

[1] The scale divisions represent category centres in this case.

Table 1 and Table 3 show the translations of the verbal anchors used for the TS and TP scales in both SCATs and Annex 69, for the Swedish and Greek language versions.

Table 2: Language type characterisation used in Annex 69, reproduced here from (Schweiker *et al.*, 2020)

Language type	group of language with respect to the number of adjectives used for the sensation scale	1: one adjective for cool and warm side each (e.g. slightly warm, warm, and very warm) 2: two adjectives for cool and warm side each (e.g. slightly warm, warm, and hot) 3c: two adjectives on the cold side and one adjective on warm side of sensation scale 3h: two adjectives on the warm side and one adjective on the cold side of sensation scale

The translations in Greek and Swedish produced in Annex 69 addressed limitations identified in (Humphreys et al, 2016) for the TS scale[2]. The application of the Greek survey in SCATs had resulted in:

a. No votes for 'hot' (+3) whilst the 'slightly warm' (+2) region was comparatively wide. One explanation for this- other than the influence of the ambient conditions- is that the scaling of the word 'ζεστά' (meaning warm or hot, depending on context) achieved using adverbs corresponding to 'slightly/very/very much' created a meaning for the equivalent of 'hot' that is not commonly experienced in the built environment.
b. Very few votes for 'cool' (-2) and of a similar number to the 'cold' votes. One possible explanation for this is that the equivalent of 'cool' (-2) used in the Greek version in SCATs would translate into 'very cool', having little difference in meaning to the phrase used for the 'cold' category (-3) that translates to 'slightly cold'.

A further observation on the Greek translation of the sensation scale in SCATs is that it has confounding affective qualities, whereas nowadays a distinction between objective and evaluative ratings is recommended (CEN, 2019):

c. The equivalent of 'neutral' translates in English as 'comfortable ideal' and has a similar leading meaning in Greek too.
d. Equally the 'slightly warm' (+1) or 'slightly cool' (-1) equivalents contain the word 'αναπαυτικά' that in English would translate as 'comfortably'. In fact the verbal anchors of these two categories are nearly equivalent to the 'comfortably warm' and 'comfortably cool' categories of the Bedford scale, which was devised for coding interview results concerning thermal (dis)comfort.

In addressing (c) and (d), the Greek version of the thermal sensation scale used in Annex 69 was designed to be free of evaluative or preferential clues. It uses the direct translation of the word neutral for (0). In addressing (b) it uses a different adjective to denote 'cold' (κρύο) from 'cool' (δροσερά) but differently to SCATs. The word-to-word translation in English is:

in SCATs: *slightly cold (-3) very cool (-2) comfortably slightly cool (-1)*
in Annex 69: *cold (-3), cool (-2), slightly cool (-1).*

In addressing (a) it uses a variation of the word 'warm' (ζεστά) to describe the three positive points of the scale. The latter is similar to the SCATs version, but with a different use of adverbs. The word-to-word translation in English is:

in SCATs: *comfortably slightly warm (+1) very warm (+2) very much warm (+3)*
in Annex 69: *slightly warm (+1), warm (+2), very warm (+3).*

[2] No publication is known to the authors discussing the other scales' translation used in SCATs.

In Annex 69 the number of adjectives used by each language version (i.e. the 'language type' in Table 1 and Table 2) was monitored so its influence on the scales' behaviour could be examined. The language type applicable to the Greek version was shared only with one other language version (Farsi) amongst the total 21 included in that survey (more in discussion). An alternative translation of 'hot' (for +3) in Greek would be '*καυτά'*. However, during the preparation of the Annex 69 translation the use of the term was reviewed in relation to point (a) but rejected, as it was felt that in the description of the indoor environment the term would only suit very hot surfaces that are hazardous to the touch, e.g. a radiator. No other suitable term was identified to use as equivalent of warm or hot other than 'ζεστά' in this context.

It is worth noting here that in Annex 69 all translations had to follow a uniform approach to questioning and presentation of the scales. An alternative translation in Greek could have employed a verb that encompasses the meaning of "feeling warm" ('ζεσταίνομαι) and another for "feeling cold" ('κρυώνω'), and create scalable meanings using adverbs. This version would correspond better to everyday language used to express thermal sensation in Greek. It would however result to a communication of thermal sensation that would deviate from the other language versions, with potentially negative implications for data analysis.

The Swedish version in the SCATs survey had fewer limitations compared to the Greek version, but some issues were identified:

a. The word 'sval' used in (-2), despite it being a word-to-word translation of 'cool', has a generally positive meaning ('fresh'). Its use in the scale may result to evaluative ratings. This is interesting as it shows that even if a translation is the closest possible to the base language (in this case English), it may not provide a suitably equivalent meaning.
b. 'Het', although the direct translation of 'hot', is not typically used in this context in everyday life. Similar to the Greek word 'καυτά', the word 'het' would be used to describe, for example, a hot object.

In Annex 69 it was decided to replace the leading phrasing for 'cool' and the unusual wording for 'hot' to address these issues. The word used for 'warm' (varm) was combined with an adverb equivalent to 'very' (mycket) to represent 'hot'. The scale produced uses one adjective for the cool and warm side each (language type 1 in Table 2) and an adverb meaning 'slightly' (något) in points -1 and +1. In the SCATs version, the direct translation of 'neutral' was used for 0. In Annex 69 'neutral' is replaced by 'neither warm nor cold', as proposed by ISO10551 for language type 1.

Table 3: Translations of verbal anchors for thermal preference scale in SCAT and Annex 69.

English (SCAT) 5pt	**English (Annex 69) 7pt**	**Greek (SCAT) 5pt**	**Greek (Annex 69) 7pt**	**Swedish (SCAT) 5pt**	**Swedish (Annex 69) 7pt**
-	much cooler	-	Πολύ πιο δροσερά	-	Mycket kallare
much cooler	cooler	*	Πιο δροσερά	Mycket varmare	Kallare
a bit cooler	slightly cooler	*	Λίγο πιο δροσερά	Lite varmare	Något kallare
without change	without	*	Καμιά αλλαγή	Ingen	Varken kallare

	change			förändring	eller varmare
a bit warmer	slightly warmer	*	Λίγο πιο ζεστά	Lite svalare	Något varmare
much warmer	warmer	*	Πιο ζεστά	Mycket svalare	Varmare
-	much warmer	-	Πολύ πιο ζεστά	-	Mycket varmare

**The Greek SCAT translation of the TP scale was not traced.*

2.3. SCATs survey context

The SCATs sample population were office workers recruited on a voluntary basis, as the project required their involvement in a yearlong study with monthly interruptions (Stoops, 2001).

The SCATs transverse, monthly surveys in Greece took place in the period September 1998 to September 1999 but no data was obtained between November 1998 and January 1999 and August 1999. Data was collected from 5 different office buildings in Athens, with two being naturally ventilated (heating in winter), one being mixed mode (heating in winter, cooling when needed in summer) and two centrally air-conditioned (heating and cooling). In total 325 subjective votes for TS and TP (each) were collected. The measured indoor air temperature during data collection ranged between 19.1°C to 30.5°C.

The SCATs transverse, monthly surveys in Sweden took place in the period August 1998 to September 1999 but no data was obtained for January and July 1999. The data was collected from 5 different office buildings located in Malmo, Gothenburg and Halmstad, with four being centrally air conditioned (heating and cooling) and one using air-conditioning for heating and partial cooling. In total 970 subjective votes for TS and TP (each) were collected. The measured indoor air temperature during data collection ranged between 18.8°C to 26°C.

2.4. Annex 69 survey context

All surveys conducted within Annex 69 involved University students who had no prior specialist knowledge in thermal comfort. Each student participated only once.

The surveys conducted in Athens took place at the National Technical University of Athens, in October 2018 (autumn) and December 2018 (winter). Participants were undergraduate students. All students were native Greek speakers (including Greek Cypriots). The questionnaires were distributed in the form of hard copies in naturally ventilated, free-running lecture theatres in both seasons. The measurement of internal air temperature at the time of the survey was 24°C in autumn and 19°C in winter. The distribution of TS responses varied between seasons but the conditions experienced were considered by the majority as comfortable or slightly comfortable and clearly or just acceptable.

The surveys in Gothenburg were conducted at Chalmers University of Technology in May 2017 and January/February 2018. All participants were undergraduate students. Four surveys were conducted, two in each season, and all took place in mechanically ventilated, lecture halls. The measurement of internal air temperatures at the time of the survey was 22.5°C and 21.8°C in spring and 20.4°C and 21.9°C in winter. Most of the respondents felt either neutral or slightly warm, which was considered by the majority comfortable and acceptable.

3. Results and discussion

3.1. Stability of subjective scale behaviour in Annex 69 and SCATs

To check the stability of the scale versions analysed here, a comparison of the psychological category widths of the same scale used in different occasions within each project is performed, as in (Humphreys et al, 2016). For Annex 69 the distinction is made between the seasons studied. For SCATs the distinction is made between the first and the second half of the year-round survey.

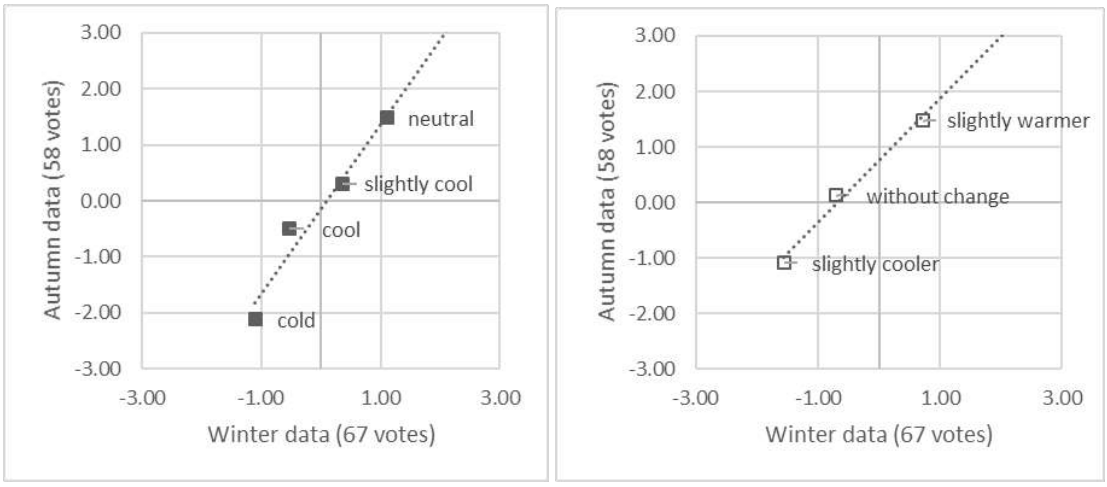

Figure 1: TS (left) and TP (right) scales' stability in Annex 69 for Athens autumn and winter surveys. The graphs show comparisons of category upper boundaries between the two seasons (correlation is high in both cases with r=0.977 for TS and r= 0.993 for TP).

A stability check for the TS and TP scales in Greek used in Annex 69 is done by applying the method of successive categories to the data from each season (winter and autumn) separately. The comparison depicted in Figure 1 (TS at the left hand side, TP at the right hand side of the figure) shows that there is a high correlation between the two sets of category boundaries for both scales (r=0.977 for TS and r=0.993 for TP). This indicates that the TS and TP scale categories had the same meaning in both seasons and therefore the scale had a stable behaviour[3] across different thermal conditions. A similar test is done for the SCATs data and the results in Figure 2 show consistency in the way the scales were interpreted by the survey participants in the two batches of data.

[3] It is observed that the gradient is not unity, as noted also in Humphreys et al (2016) and attributed to the different environmental conditions present.

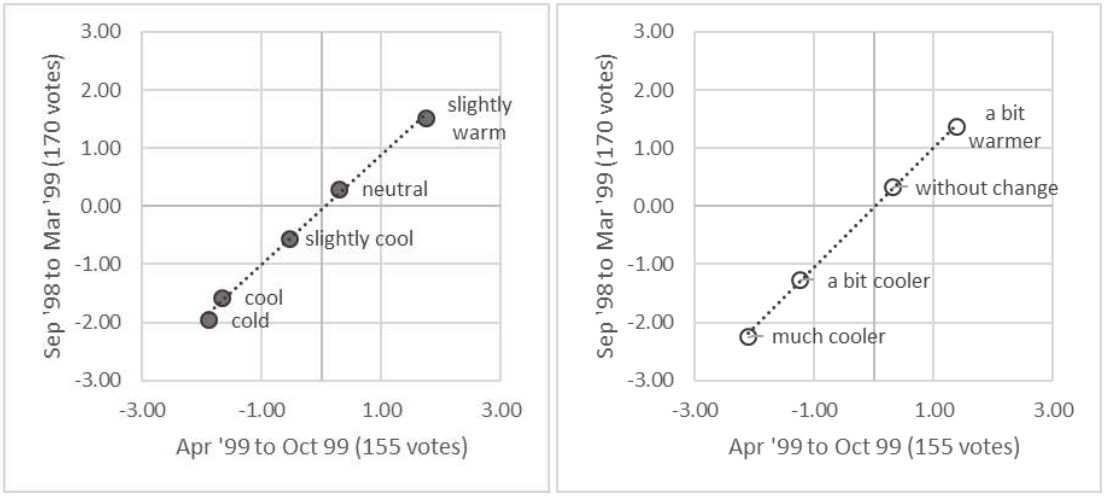

Figure 2: TS (left) and TP (right) scales stability in SCATs, first and second half of year-round survey done in Greece (Athens). The graphs show comparisons of upper category boundaries between the two batches of data (correlation is high in both cases with r=0.998 and r=0.999 respectively).

The results for the Swedish data show similar consistency in the relative category widths as seen in Figure 3 and Figure 4. Following these checks the combined data from each survey (data from both seasons in Annex 69 and year-round data for SCATs) is used in the analyses below, even though the thermal environments represented in the combined data are not always similar. This is in line with how the method was previously applied (Humphreys et al, 2016). It is observed that some category upper boundaries are missing in Figures 1-4. This is due to lack of data for that category in either or both seasons (no Probit value).

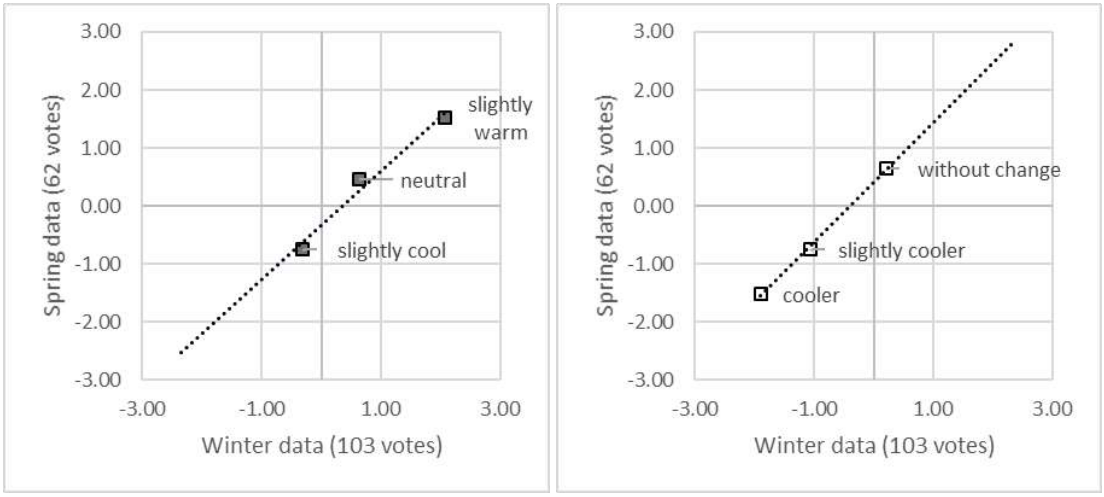

Figure 3: TS (left) and TP (right) scales stability in Annex 69, Gothenburg spring and winter surveys. The graphs show comparisons of upper category boundaries between the two surveys done in spring and winter (correlation is high in both cases with r=0.988 for TS and r= 0.999 for TP).

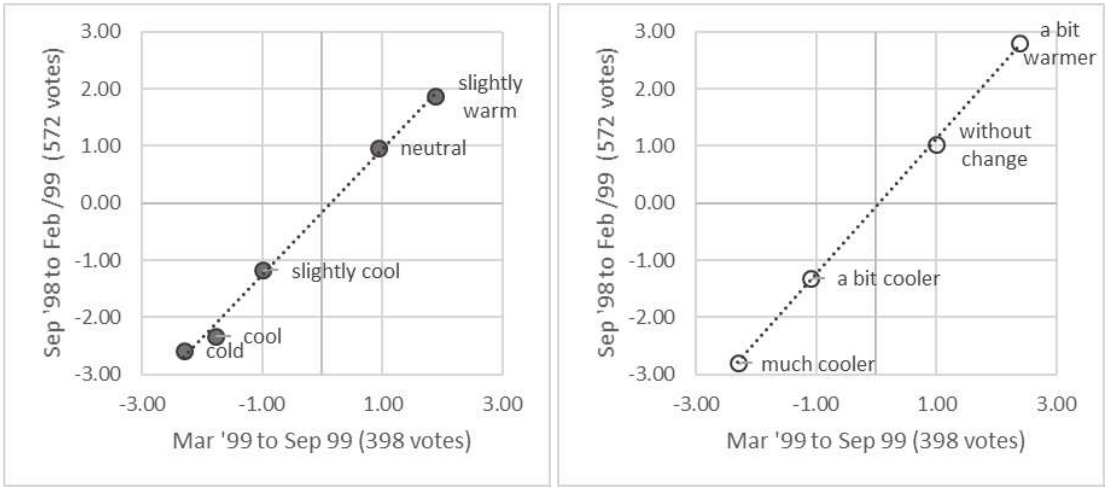

Figure 4: TS (left) and TP (right) scales stability in SCATs, first and second half of year-round survey done in Sweden (Malmo, Goteborg and Halmstad). The graphs show comparisons of upper category boundaries between the two batches of data (correlation is high in both cases with r=0.997 for TS and r= 0.999 for TP).

3.2. Category widths and scales' characteristics

A comparison between the category widths resulting from the Greek questionnaires used in SCATs and Annex 69 is shown in Figure 5, along with calculated Standard Error (SE) in probit units[4].The method used does not allow for direct comparisons between different versions of scales, but a visual comparison is useful as it shows the behaviour of each version and the accuracy of the category widths presented.

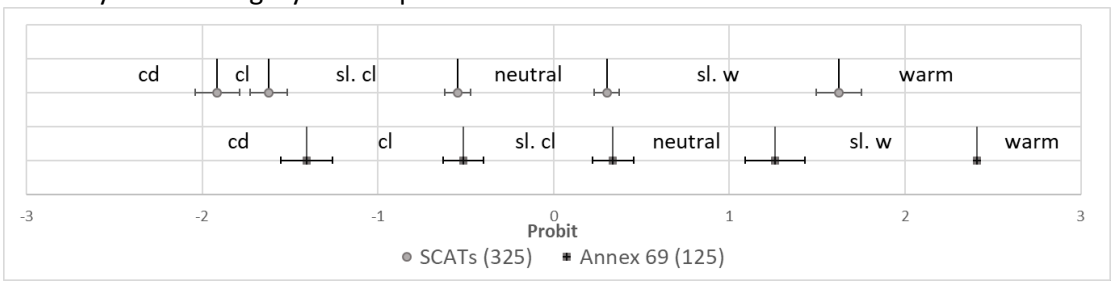

Figure 5: TS category widths for the two scales used in the questionnaires in Greek in SCATs and Annex 69 and standard error in probit units (not shown for upper boundary of slightly warm in Annex 69).
Wording used is from the equivalent English translation given in Table 1.
Legend: cd: cold, cl: cool, sl. cl.: slightly cool, sl. w: slightly warm

The SE is higher in Annex 69 due to the smaller sample (125 votes) in comparison to SCATs (325 votes). The SE is also higher at the margins of the scale, due to small differences in proportions resulting to large differences in Probit units. The SE of the upper boundary of the slightly warm category in the Annex 69 scale in Greek (SE=1.604) is not depicted as it is so large that it would conceal the category boundaries of the slighty warm and neutral categories. The accuracy levels presented in this figure indicate that for both sets of surveys we may draw reliable conclusions for the centre of the data, but any observations made for the margins need to be interpreted with caution.

The properties of the two scales appear dissimilar, with the scale used in Annex 69 producing a better distribution of category widths on the continuum. It appears that the

[4] As in (Humphreys et al., 2016), with SE of a binomial proportion being $\sqrt{(\frac{p(1-p)}{n})}$, where *p* is the cumulative proportion and *n* is the total number of responses in that dataset.

rewording of the categories, using the rationale discussed previously, has addressed the issues encountered in SCATs with the category widths of the cool side of the scale. However, similarly with the SCATs survey, none of the respondents in the Greek surveys of Annex 69 felt 'hot', hence this category is not depicted in Figure 5. In both Annex 69 and SCATs the 'slightly warm' category is larger than any other, however the boundaries are subject to a large SE and therefore of questionable accuracy. It is likely that as the data from Annex 69 is not numerous and the surveys took place in autumn and winter, the lack of responses for 'hot' in this dataset is due to the thermal conditions experienced (24°C in autumn and 19°C in winter). Indeed the median respondent in this case (the zero point on the Probit scale) is found in the slightly cool category. In the SCATs data the middle of the neutral category falls near the zero point of the psychological continuum, reflecting on the sensation of the median participant in the office buildings surveyed. As the SCATs data for Greece is more substantial in size than in Annex 69 and it includes subjective responses from an annual survey that was more detailed in summer than in winter, the absence of 'hot' votes stands out and may indeed be due to the wording choices made. However this observation needs to be further examined in future studies using sufficiently large samples for analysis together with a larger variation of prevailing thermal indoor conditions. Another possible reason for the behaviour of the warm side of the scale can be deduced from the outcomes presented in (Schweiker et al., 2020). The Annex 69 study included a set of free-positioning tasks for establishing the perceived widths of categories in each scale and the relationships between scales. The existence of participants' subgroups sharing common perceptions with regards to the scales was examined. It was found that there was a significant difference due to language type in the distribution of subgroups for the free-positioning task for TS anchors on the thermal comfort scale, with the language type 3c (applying to the Greek translation) having the highest impact amongst all language types on the respondents' interpretation of the scales. This seems to indicate that the use of two adjectives on the cold side and one adjective on warm side of sensation scale results to an interpretation of the scale that is distinct to other language types and needs to be revisited.

A comparison between the category widths resulting from the Swedish questionnaires used in SCATs and Annex 69 is shown in Figure 6, along with calculated Standard Error (SE) in probit units. As with the Greek data, a visual comparison is made to examine the behaviour of each version and the accuracy of the category widths. The boundaries between the three middle categories are more precisely defined in comparison to the same boundaries resulting from the Greek data. This is due to the larger samples available from the Swedish surveys. As with the Greek data, the reliability of any conclusions drawn for the margins of the data is low.

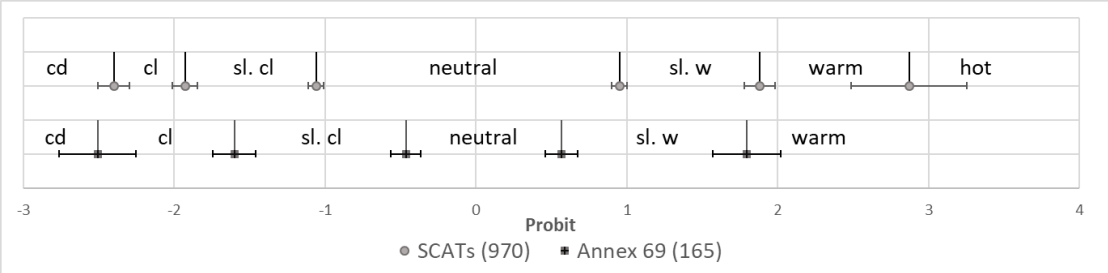

Figure 6: TS category widths for the two scales used in the questionnaires in Swedish in SCATs and Annex 69 and standard error in probit units.

Wording used is from the equivalent English translation given in Table 1.

Legend: cd: cold, cl: cool, sl. cl.: slightly cool, sl. w: slightly warm

The properties of the two scales appear dissimilar, with the scale used in Annex 69 producing a better distribution of category widths on the psychological continuum. The rewording of the neutral category appears to have significantly contributed to this improvement, indicating that in some languages the recommendation made by the relevant standard is useful or that in general, the usage of "neither cold nor warm" should be preferred compared to "neutral". It also appears that the rewording of the categories, using the rationale discussed previously, has addressed the issues encountered in SCATs with the category widths of the cool side of the scale. However, none of the respondents in the Swedish surveys of Annex 69 felt 'hot' ('mycket varm') due to the moderate thermal conditions experienced during the surveys (22.5°C and 21.8°C in spring and 20.4°C and 21.9°C in winter), hence this category is not depicted in Figure 6.

The middle of the neutral category in both Swedish scales falls near the zero point of the psychological continuum, reflecting on the sensation of the median participant in the buildings surveyed. In SCATs this position is slighlty offset to the cool side, whilst in Annex 69 is offest to the warm side. These differences are however comparable to the SE shown, and are potentially insignificant to any conclusions drawn.

3.3. Comparing the behaviour of TSV and TPV scales

Humphreys et al (2016) presented a comparison between TS and TP scales within the same body of survey data, using an example that applied the method of successive categories on data collected using the ASHRAE (sensation) and the Nicol preference scales. The same process is applied here to examine the relationship between the TS and TP scales used in SCATs, separately for the Greek and the Swedish surveys. This is a straightforward comparison in the SCATs questionnaire, given that the order of categories in the two scales are in agreement:

TS: How do you feel at this time:
Cold/ Cool/ Slightly cool/ Neutral/ Slightly warm/ Warm /Hot

TP: I would prefer to be:
Much warmer/A bit warmer/No change/A bit cooler/Much cooler

However performing the same for the TS and TP scales in Annex 69 requires some data adjustment because the TP scale was presented to participants in the reverse order to the Nicol preference scale:

TS: How do you feel right now:
Cold/ Cool/ Slightly cool/ Neutral/ Slightly warm/ Warm /Hot

TP: At this moment, would you prefer to be:
Much Cooler/Cooler/ Slightly cooler/ Without change / Slightly warmer/Warmer/ Much warmer

The rearrangement of the data involves recoding the scale so that category 1 (-2) is Much warmer and category 5 (+2) is Much cooler. The cumulative proportions and Probits for TP are then calculated for this new order. This manipulation assumes that the presentation of the scale in the reverse order has no effect on the scale behaviour and on the notion of the psychological continuum.

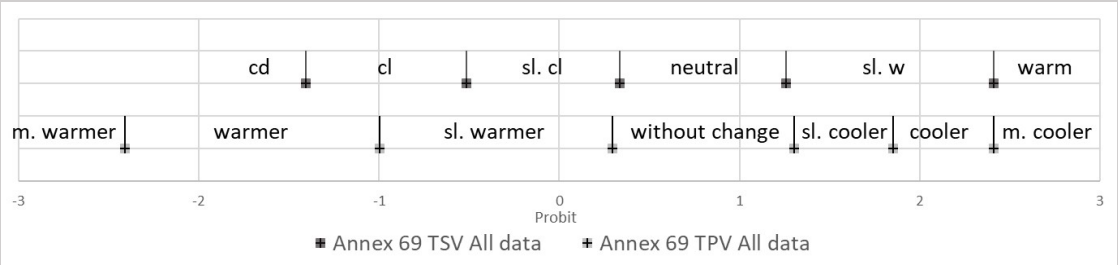

Figure 7: Relationship between TSV and TPV categories for Greek survey data, Annex 69.
Wording used is from the equivalent English translations given in Tables 1 and 3.
Legend: cd: cold, cl: cool, sl. cl.: slightly cool, sl. w: slightly warm, m.: much, sl.:slightly (warmer/cooler)

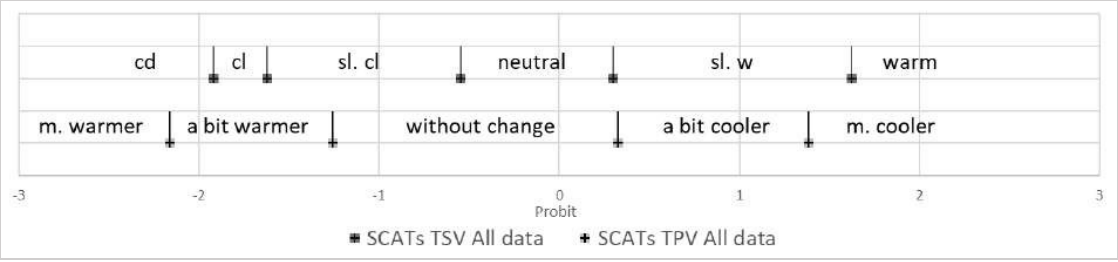

Figure 8: Relationship between TSV and TPV categories for Greek survey data, SCATs.
Wording used is from the equivalent English translations given in Tables 1 and 3.
Legend: cd: cold, cl: cool, sl. cl.: slightly cool, sl. w: slightly warm, m.: much (warmer/cooler)

The 'neutral' and 'without change' categories of the Greek translations produced in Annex 69 are more aligned, in comparison to the same categories in SCATs. In SCATs the "without change" category is quite wide, corresponding to a wider part of the psychological continuum. This may be due to the leading meaning given in the -1 and +1 categories in SCATs for this language version, as previously discussed. More specifically, the data analysis shown in Figure 8 may indicate that in this context most of the respondents who felt 'comfortably cool' (in those conditions), may in fact thought that this was the ideal environment and preferred no change. The fact that the upper boundaries of the two categories appears to be in agreement may indicate that the sensation of 'comfortably warm' is less equivalent of the ideal in comparison to the 'comfortably cool', corresponding to a preference for 'a bit cooler'. No further conclusions are drawn from this comparison for the margins of the data, as the expected accuracy (as previously discussed) is low.

A similar set of analyses is performed for the Swedish versions of the TS and TP scales. Figure 9 shows that there is a strong agreement between the TS and TP scales in Annex 69 for this language version and in particular for the cool side of the scales. There is less agreement in the warmer side of the scales, starting with the upper boundary of the 'without change' category falling within the region of the 'slightly warm' sensation. This may indicate that in a cold context like this, some of the respondents who feel 'slightly warm' may find this environment ideal and prefer no change. The existence of votes for a 'much cooler' preference may be seen as evidence that the wording chosen for the +3 category of TS was unsuccessful or that a 'warm' sensation is experienced more intensely in a cold context. Figure 10 shows the relationship between the TS and TP scales in Swedish used in SCATs. In this case, there is relatively good agreement between the 'neutral' and 'without change' categories, which could be due to the use of the word "neutral" in this version. People's interpretation of the word "neutral" in relation to thermal sensation and its equivalence to "neither cool nor warm" is a matter for further investigation as presented by

Schakib-Ekbatan et al. (2018) for the German context. No other patterns are observed with regards to the boundaries of the other categories in Figure 10.

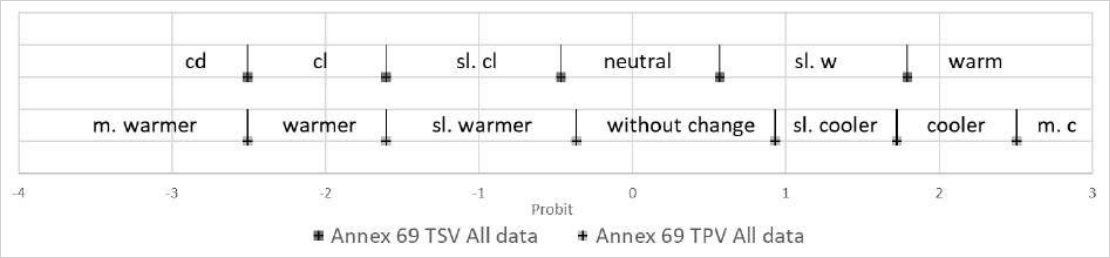

Figure 9: Relationship between TSV and TPV categories for Swedish survey data, Annex 69.

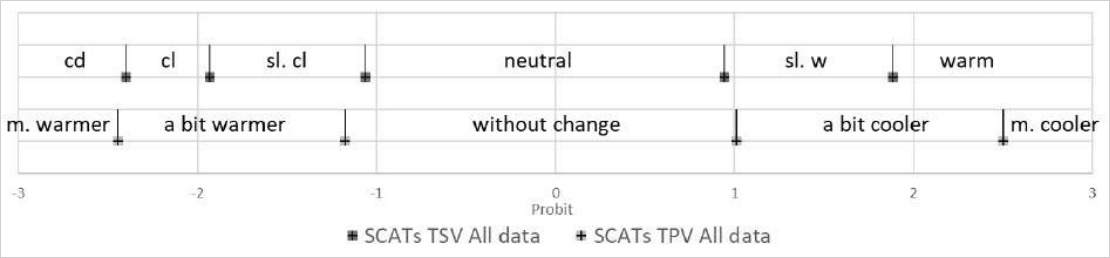

Figure 10: Relationship between TSV and TPV categories for Swedish survey data, SCATs

4. Conclusions

The TS scales produced in Annex 69 for the Greek and Swedish questionnaires addressed limitations previously observed with the equivalent scales used in SCATs. These revised scale translations were found to produce a better distribution of category widths on the psychological continuum, in particular with regards to irregularities previously noted with the cool side of thermal sensation in both language versions. As the marginal category boundaries are subject to a large SE (as resulting from the use of small survey samples), it is impossible to deduce whether the lack of any votes for 'hot' in both language versions in Annex 69 is due to inappropriate wording choices or due to the prevailing cool survey conditions. However some suggestions for further improving these translations were presented in the discussion above.

Juxtposing the category widths of the TS and TP scales reveals that the assumed equivalence between neutrality and preference for no change may not exist, especially when the TS and TP scales are designed with the intention to separate objective and affective ratings within a survey; in the cold Swedish climate, respondents may rate the environment as warm but ideal, whilst in the warm, fuel-poor Greek context a "comfortably cool" sensation incorporates a preferential clue for 'no change'. The findings presented here indicate that if conclusions are to be drawn for a particular environment from the subjective votes collected using more than one scale, the vertical association of terms (between the related categories across these scales) needs to be carefully considered as much as the horizontal (between categories within the same scale). Equally important is the correct alignment (of ranking order) of the various scales considered, as any rearrangements made retrospectively may introduce error to the analyses. In fact future work needs to revisit the assumption made here, that the reversing of the TP Annex 69 data has no impact on the resulting scale properties, as recent observations on the distinct behaviour of right-to-left languages may suggest that this is not always true (Schweiker et al., 2020).

5. Acknowledgements

The authors would like to thank Michael Humphreys for the advice offered in relation to the method of successive categories and the reordering of the thermal preference scale. They are also grateful to Aris Tsagrasoulis, Fergus Nicol and John Stoops for sharing information and data from the SCATs project.

This work has been performed within the framework of the International Energy Agency - Energy in Buildings and Communities Program (IEA-EBC) Annex69 "Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings". www.iea-ebc.org, www.annex69.org. The preparation of the Annex 69 questionnaire versions was a collaborative effort and therefore thanks are extended to the entire team (for a complete list of contributors see the relevant cited publication).

6. References

Al-Khatri, H. and Gadi, M. B. (2019) 'Investigating the behaviour of ASHRAE, Bedford, and Nicol thermal scales when translated into the Arabic language', *Building and Environment*, 151, pp. 348–355.

CEN (2019) 'BS EN ISO 10551:2019 *Ergonomics of the physical environment. Subjective judgement scales for assessing physical environments*'.

Guilford, J. P. (1954) *Psychometric methods*, 2nd ed., New York: McGraw-Hill.

Humphreys, M., Nicol, F. and S., R. (2016) *Adaptive Thermal Comfort: Foundations and Analysis*. Abingdon: Earthscan.

Khatun, A. et al. (2017) 'Differences in reported linguistic thermal sensation between Bangla and Japanese speakers', *Journal of Physiological Anthropology*. BioMed Central Ltd., 36(1).

Ličina, V. F. et al. (2018) 'Development of the ASHRAE Global Thermal Comfort Database II', *Building and Environment*, 142, pp. 502–512.

McCartney, K. and Nicol, F. (2002) 'Developing an adaptive control algorithm for Europe', *Energy and Buildings*, 34, pp. 623–635.

Nicol, J. F. and McCartney, K. (2000) SCATs. *Final report for Task 2: Conducting comfort surveys throughout Europe*.

Pitts, A. (2020) 'The Languages and Semantics of Thermal Comfort'. in: *NCEUB Conference: Comfort and Energy Use in Buildings- Getting them right*, Cumberland Lodge, Windsor.

Rijal, H. B. _et_a. (2012) 'A comparison of the winter thermal comfort of floor heating systems and air conditioning systems', in: *5th International Building Physics Conference*. Kyoto, Japan.

Schakib-Ekbatan, K., Becker, S., Cannistraro, A., & Schweiker, M. (2018). What do people associate with "cold" or "hot"? - Qualitative analyses of the ASHRAE-scales' labels. In *10th Windsor Conference - Rethinking Comfort* (pp. 371–386).

Schweiker, M. et al. (2019) 'The Scales Project, a cross-national dataset on the interpretation of thermal perception scales', *Scientific data*, 6(1).

Schweiker, M. et al. (2020) 'Evaluating assumptions of scales for subjective assessment of thermal environments – do laypersons perceive them the way, we researchers believe?', *Energy and Buildings*, 211.

Stoops J. (2001) The physical environment and occupant thermal perceptions in office buildings: an evaluation of sampled data from five European countries. Göteborg, Sweden: Chalmers University of Technology.

Williamson, T. and Daniel, L. (2018) 'Changing thermal comfort expectations: studies in Darwin, Australia.', in: *10th Windsor Conference: Rethinking Thermal Comfort*, April 2018, pp. 68–86.