The linearly weighted Kappa

Interrater reliability is the extent to which two or more individuals (coders or raters) agree. Various coefficients of agreement are available to calculate interrater reliability. At least ordinal level of measurement was presumed for the items of the COMFORT scale, which consist of five closed response categories. Therefore we used the linearly weighted Cohen's Kappa (Cohen 1968). The interested reader may refer to the method of calculation for this index. Otherwise, several statistical programs are available to calculate the weighted Cohen's Kappa. For instance PROC FREQ in SAS, or AGREE, a software program produced by *Pro*GAMMA (http://www.gamma.rug.nl/)
In general, the weighted Kappa is used if you want to weight the close misses (e.g., 1-2) more heavily than misses that are further apart (e.g., 1-3).
The weighted Kappa makes sense if the rating scale is on an ordinal scale where a 1 point difference could be considered less of an incorrect rating than a 2 point difference. Although typically the weighted Kappa is a higher estimate than simple Kappa, this is not necessarily true if there are more larger differences (e.g., 1-3) than smaller differences (e.g., 1-2). Weighted kappa is typically not appropriate for purely categorical variables where there is no ordering of the values.

Formula for the linearly weighted Kappa:

$\kappa_w = (P_o - P_e)/(1-P_e)$

in which P_o is the proportion weighted observed agreement,
and P_e is the proportion weighted chance agreement.

The linearly weighted Kappa is determined by a specific weight matrix in which each weight is calculated by the following rule:

$W_{ij}=1-\{|i-j|/|c-1|\}$ with c being the total number of response categories (=5 in the COMFORT scale).

the weight matrix in this case is:

	1	**2**	**3**	**4**	**5**
1	1.00	0.75	0.50	0.25	0.00
2	0.75	1.00	0.75	0.50	0.25
3	0.50	0.75	1.00	0.75	0.50
4	0.25	0.50	0.75	1.00	0.75
5	0.00	0.25	0.50	0.75	1.00

Example:

		Scoring of rater 1 on 'facial tension'				
Scoring of other rater(s) on 'facial tension'		**1**	**2**	**3**	**4**	**5**
	1	**8** *(0.07)*	**2** *(0.02)*	**1** *(0.009)*		
	2	**3** *(0.026)*	**11** *(0.09)*	**5** *(0.04)*		
	3		**7** *(0.06)*	**55** *(0.47)*	**11** *(0.09)*	
	4			**1** *(0.009)*	**11** *(0.09)*	
	5					**2** *(0.02)*
		Total	**117 pairwise ratings**			

Between brackets and italics: proportion of occurrence

P_0= 0.07*1+0.02*0.75+0.009*0.50+ 0.026*0.75+0.09*1+0.04*0.75+
0.06*0.75+0.47*1+0.09*0.75+0.009*0.75+0.09*1+
0.02*1≈0.93

To calculate P_e the proportions of the margins are required

	1	**2**	**3**	**4**	**5**	
1	**8** *(0.07)*	**2** *(0.02)*	**1** *(0.009)*			**0.0940**
2	**3** *(0.026)*	**11** *(0.09)*	**5** *(0.04)*			**0.1623**
3		**7** *(0.06)*	**55** *(0.47)*	**11** *(0.09)*		**0.6239**
4			**1** *(0.009)*	**11** *(0.09)*		**0.1026**
5					**2** *(0.02)*	**0.0171**
	0.0940	**0.1709**	**0.530**	**0.1880**	**0.0171**	

P_{e}=1.00*(0.094*0.094)+0.75*(.094*0.1709)+0.50*(0.094*0.530)+0.25*(0.094*0.118)+0+

0.75*(0.1623*.094)+1.00(*0.1623*0.1709)+0.75*(0.1623*0.53)+0.50*(0.1623*0.188)+0.25*(0.1623*0.0171)+

0.50*(0.6239*0.094)+0.75*(0.6239*0.1709)+1.00*(0.6239*0.53)+0.75*(0.6239*0.188)+0.50*0.6239*0.0171)+

0.25*(0.1026*0.094)+ 0.50*(0.1026*0.1709)+0.75*(0.1026*0.53)+1.00*(0.1026*0.188)+0.75*(0.1026*0.0171)+

0.25*(0.0171*0.1709)+0.50*(0.0171*0.53)+0.75*(0.0171*0.188)+1.00*(0.0171*0.0171) ≈0.78

The linearily weighted Kappa = 0.6924 = (0.9338 - 0.7846) / (1.0000 - 0.7846)