Measuring mood – relative sensitivity of numerical rating and Likert scales in the context of teaching electroacupuncture. Initial findings and the influence of response style on results.

© David Mayor,[1] Tony Steffert[2]
1. University of Hertfordshire; 2. Open University

Poster presented at the 18th International Acupuncture Research Symposium,
King's College London, 19th March 2016

CONTENTS

[1] davidmayor@welwynacupuncture.co.uk.
[2] tony@qeeg.co.uk

ABSTRACT

Background. Mood and its changes are important but often overlooked in acupuncture research. A number of mood questionnaires exist, all using Likert scales, such as the validated, open-access 24-item Brunel Mood Scale ($BRUMS_{24}$). This comprises six subscales – five of negative valence (ANGer, CONFusion, DEPRession, FATigue, TENSion), one positive (VIGour), with a derived construct, Total Mood Disturbance (TMD).

Our own acupuncture research called for a quick and easy method of assessing mood and its changes before, during and after stimulation. Following several pilot studies, we developed a multiple numerical rating scale for mood (NRS-M) with four negative subscales (Anxious, Confused, Fatigued, Gloomy) corresponding to those in $BRUMS_{24}$, and four positive subjective states (Comfortable, Lively, Relaxed and overall 'Good mood') that would be useful in our research, rather than focusing solely on the pathological states emphasised by $BRUMS_{24}$. Using fewer words, we hoped this would be easier to understand for non-native English speakers and those with learning difficulties, as well as quicker to complete even than $BRUMS_{24}$. Here we analyse its use in a teaching situation.

Main objectives. To assess validity and reliability of NRS-M and determine whether it is more sensitive to mood and its changes than $BRUMS_{24}$. To conduct factor and cluster analyses for NRS-M and $BRUMS_{24}$, and to investigate common response styles for both scales.

Methods. Ethics approval was obtained, and 158 respondents recruited from attendees at nine UK electroacupuncture (EA) training seminars. The two measures were administered in paper form in quick succession – once early on and once towards the end of the seminar. Non-parametric statistical methods were used.

Results. NRS-M concurrent validity with $BRUMS_{24}$ was good, with better discriminant validity for NRS-M, but better internal consistency and test-retest reliability for $BRUMS_{24}$ (suggesting greater sensitivity for NRS-M). Of six other methods used, four indicated that NRS-M was more sensitive to mood and its changes than $BRUMS_{24}$. Significance levels for Pre-Post differences in corresponding subscales in NRS-M and $BRUMS_{24}$ were very similar, but effect sizes were all small. Results of factor and cluster analysis were consistent for NRS-M, showing two main factors/clusters, one of positive valence and one negative. Most NRS-M respondents exhibited a 'mild' response style, RS (or non-extreme RS, NERS). Thus response style (RS) is unlikely to threaten the validity of conclusions drawn from responses to the two scales. However, ERS respondents to NRS-M are likely to emphasise positive moods and de-emphasise negative moods ($p<0.001$).

Conclusions. NRS-M is a reasonably robust, responsive and partially validated measure of mood to use in the context of acupuncture practice. It has advantages over longer Likert scales such as $BRUMS_{24}$.

Background

What is mood?

Acupuncturists trained in the Five-Element tradition often assess patients in terms of the five emotions – Joy, Sympathy, Grief, Fear and Anger. These are not unrelated to the four biologically-based primary emotions and their opposites postulated by Plutchik (Plutchik 1958), popularised in his 'wheel of emotions' and assessed using his Emotions Profile Index (Plutchik & Kellerman 1984). However, 'emotions' are not necessarily the same as 'moods'.

Mood and emotion have been described as closely related but distinct constructs (Beedie et al. 2005). Both involve value judgment based in feeling (Whitmont 1978), but mood is generally taken to be more nebulous and pervasive, dependent more on our subterranean inner tendencies, or what Carl Jung called our unconscious – or only partially conscious – 'complexes' (Jacobi 1959; Whitmont 1978). Emotion, in contrast, may be more overt and easily identifiable, and is often a reaction to an immediate outward event or situation (Alston 1967). Mood can thus be seen as more a pervasive state of mind or emotional 'climate', emotion more like the changeability of 'weather' (Nowlis 1970). They differ according to a number of criteria – in particular duration (mood may be longer lasting), cause (emotion is usually more clearly defined) and intentionality (emotion being about something more specific, whereas mood may not be 'about' anything in particular) (Beedie et al. 2005). However, those who have developed measures of mood rarely discuss such distinctions, and the terms 'mood' and 'emotion' are often used interchangeably (Lorr 1989; Plutchik 1989) – even by those who have contributed to the debate on their differences (Terry et al. 2003; Lane et al. 2005).

Measuring mood

Like emotions (Plutchik 1989), moods are subjective, personal and idiosyncratic, and therefore in their very nature difficult to define and measure. Nonetheless, this has done little to dissuade people from devising a number of ways of assessing mood, whether as a trait (longer-term pattern), or as a state (more temporary).

Measures of mood and its changes are potentially important but often overlooked in acupuncture research. Mood states before treatment may affect outcome, and changes in mood may reflect other benefits of treatment – or, indeed, its adverse effects. However, PubMed searches (14 Feb 2016) revealed only eight studies with both 'acupuncture' and 'mood' in the title (and a further 12 with 'emotion*' or 'feeling'). Of these, only two studies used composite scales that assessed more than one dimension of mood and were applicable in healthy persons (Acker et al. 2015; Sawazaki et al. 2008).

A number of mood questionnaires exist, with early versions reviewed by Lorr (1989). Virtually all that are still current use Likert scales, as for example the PANAS (Positive Affect and Negative Affect Scale), which considers positive and negative affect as the dominant dimensions of emotional experience (Watson et al. 1988). This was later expanded by Watson and Clark as the 60-item PANAS-X, with 23 subscales for negative emotions, 18 for positive emotions, and 14 for other affective states, such as Fatigue (sleepy, tired, sluggish, drowsy) and Serenity (calm, relaxed, at ease). Scoring is with a 5-point Likert scale (Watson & Clark 1994). At around the same time, Mayer

and Gaschke developed a simpler 16-item Brief Mood Introspection Scale (BMIS) with a 4-point Likert scale for eight positive and eight negative affects (such as Calm, Content or Lively, and Gloomy, Nervous or Drowsy), together with a 21-point numerical rating scale (NRS) for Overall mood, from 'Very Unpleasant' to 'Very Pleasant' (Mayer & Gaschke 1988). Apart from the binary Mood Profile Index (Plutchik et al. 1968), the oldest – and longest – mood assessment tool still in use is the proprietary 65-item Profile of Mood States (POMS), again based on a 5-point Likert scale (McNair et al. 1971). Shorter variants include the 30-item 'Brief' form, POMS-B (McNair et al. 1992) a 37-item 'Short' form, the POMS-SF (Shacham 1983), and even an 11-item version for use with cancer patients (Cella et al. 1987). However, these all use a number of American English terms, which in our experience of the POMS-SF some UK respondents and particularly non-native English speakers find difficult to interpret. An anglicised and shorter version of the POMS, still based on a 5-point Likert scale, is the 24-item Brunel Mood Scale ($BRUMS_{24}$), which has been validated in a number of studies (Terry et al. 1999; Terry et al. 2003; Lane et al. 2005), including a Chinese version used in a classroom setting (Zhang et al. 2014). Both the POMS-SF and $BRUMS_{24}$ comprise six subscales (ANGer, CONFusion, DEPRession, FATigue, TENSion and VIGour), and have been used in our own previous research (Mayor & Steffert 2013; Mayor et al. 2015). In contrast to the BMIS and the original PANAS, where negative and positive affects are balanced, the POMS and $BRUMS_{24}$ focus on negative moods and a Total Mood Disturbance (TMD), with only one positive subscale (VIGour).

Mood-related states

In addition to scales designed to assess mood, others have been developed to measure subjective states such as Boredom, Comfort and Relaxation. 'Boredom' as a state (rather than a trait) has received little attention, existing scales being all Likert-based (Vodanovich & Watt 2016). 'Comfort' – within the nursing context – has been defined by Kolcaba (1991) as 'the state of having met basic human needs for ease, relief and transcendence', with 'ease' itself as 'a state of calm or contentment', whether in a physical, social, environmental or psychospiritual context. In her own research, Kolcaba has used both Likert and visual analogue scales (VAS) for 'Comfortable' and its components, and demonstrated that the concurrent validity of the latter with the more established Likert version was reasonable, although the Likert scale performed better at detecting differences over time and between respondents (Kolcaba & Steiner 2000).

There appear to be few self-rating scales for Relaxation. Discounting the observational Behavioral Relaxation Scale (Schilling & Poppen 1983), self-rating measures of Relaxation include the 38-item Smith Relaxation States Inventory, which assesses 18 (more recently 19) relaxation and three stress states, using 4- or 6-point Likert scales, and with underlying factors such as sleepiness, disengagement, physical relaxation, rested/refreshed and mental quiet/relaxation (Smith et al. 2000; Smith 2001), and the 5-point Likert Relaxation Inventory (Crist et al. 1989). The latter includes separate physiological and cognitive tension scales, along with a 20-item physical assessment scale for relaxation (e.g. 'I feel content', 'I feel very calm', 'I feel refreshed'). This separation of tension from relaxation is supported by Japanese research that 'relief from tension' and 'calmness and relaxation' comprise discrete subjective emotional states (Monchi & Suzuki 2000), with analysis of the S-MARE, a Japanese relaxation measure (based on Crist et al.) demonstrating the existence of three separate factors – physiological tension, psychological relaxation and anxiety (Sakakibara et al. 2014). From this brief account, there appears to be some overlap between 'Comfortable', 'Relaxed' and 'Calm', for example.

The need for a new measure of mood for use in acupuncture research

For use in our own research on the effects of acupuncture-related interventions on the brain (electroencephalography, EEG) and heart (heart rate variability, HRV), we required a quick and easy method of assessing mood and its changes before, during and after stimulation. Most of the above measures were overly complicated for our purposes. In contrast, existing multiple VASs for mood did not meet our requirements. Examples are Stern's multiple VAMS (sad, confused, afraid, happy, tired, angry, tense and energetic), designed for use with neurologically impaired patients rather than healthy adults (Stern 1997; Arruda et al. 1999; Okun et al. 2003) and two scales for use only with unipolar or bipolar depressed patients – the Internal State Scale (Bauer et al. 1991) and the multiple VAS (MVAS) for mood ratings (Ahearn et al. 1996). We were unable to access the Emotional Assessment Scale (EAS) for review (Carlson et al. 1989), but as it includes 24 subscales and emotions such as guilt and surprise, again it was unlikely to be appropriate for our research. One-dimensional scales, such as the VASM ('worst sad' to 'best happy') (Hall et al. 1996), which only assesses global mood, the VAS-A for anxiety (Abend et al. 2014) and the GA-VAS for global anxiety (Williams et al. 2010), were also deemed inappropriate, although providing useful precedents. The only other existing candidate for our use that we could find was a questionnaire created by Abele-Brehm and Brehm (1986) and used by Acker et al. (2015) in acupuncture research, but as this provided merely binary ('yes'/'no') results, has currently only been used in the original German version – and in the proposed English version (Abele-Brehm n.d.) employs some terms which translate somewhat ambiguously – it too was discounted.

Therefore, based on our own prior experience in pilot studies with a VAS for Relaxation (Mayor & Steffert 2013), we developed a multiple numerical rating scale for mood (NRS-M) which uses a very few simple words as subscale anchors (Mayor et al. 2015). This, we hoped, would be easier to understand for non-native English speakers and those with learning difficulties, as well as quicker to complete even than $BRUMS_{24}$. Furthermore, we decided to include positive subjective states such as 'Comfortable', 'Lively' and 'Relaxed' that we considered would be useful in our research, rather than focusing on the more pathological states emphasised by the POMS and $BRUMS_{24}$. Additional moods such as those in the PANAS-X and BMIS were considered but not included, either because we did not think they would be useful in our research project, or because we had used them in a separately developed scale on feelings expected and experienced during treatment (Mayor & Steffert 2013; Mayor 2014; Mayor & McClure, in preparation). We did not consider 'Boredom' as a possible mood at the time we developed NRS-M.

One of the present authors (TS) is also currently developing a 20-point Emotional Rating Scale for use in EEG research (Steffert et al. n.d.). The eight subscales employed differ from those of the POMS, and – like the 40 items in the eight-subscale Abele-Brehm and Brehm questionnaire – are derived from Russell's two-axis 'circumplex'-style model of affect (valence, or pleasantness, and arousal, or activation) (Russell 1980; Posner et al. 2005), which was based in turn on earlier research on facial expression by Schlosberg (1952).

Objectives

1. To assess criterion (concurrent) validity of NRS-M by using it in parallel with $BRUMS_{24}$.

2. To explore convergent and discriminant validity of NRS-M compared with $BRUMS_{24}$.

3. To compare internal consistency and test-retest reliability of NRS-M and $BRUMS_{24}$.

4. To determine whether NRS-M is more sensitive to mood and to changes in mood than $BRUMS_{24}$ using a variety of methods.

5. To find out whether initial state predicts response when using NRS-M and $BRUMS_{24}$.

6. To conduct factor and cluster analyses for NRS-M and $BRUMS_{24}$.

7. To investigate common response patterns and styles for both scales, NRS-M and $BRUMS_{24}$.

8. On a small sub-sample (*N*=13), to explore whether there are any significant relationships between response patterns and emotional intelligence characteristics at baseline.

9. On a second small sub-sample (*N*=16), to check whether users find it quicker and easier to use NRS-M or $BRUMS_{24}$ and to ascertain their preferences for one or the other.

Methods

Rationale for item inclusion/exclusion

NRS-M was designed to include a number of individual subscales corresponding to the six subscale moods of the established POMS-SF and $BRUMS_{24}$, as shown in **Table 1**.

Table 1. Subscales of NRS-M and $BRUMS_{24}$.

NRS-M	$BRUMS_{24}$
Anxious	Tension (TENS)
Comfortable	[no equivalent]
Confused	Confusion (CONF)
Fatigued	Fatigue (FAT)
Gloomy	Depression (DEPR)
Lively	Vigour (VIG)
Relaxed	[no equivalent]
[no equivalent]	Anger (ANG)
Overall mood (Overall)	Total mood disorder (TMD)

In addition, 'Comfortable' – although perhaps not strictly a mood as such – was included in NRS-M because the measure was first designed to be used in studies in which participants were seated for long periods wearing an EEG (electroencephalography) electrode cap and various other

electrophysiological sensors and asked to move (and even blink) as little as possible. We were concerned that any discomfort experienced could potentially act as a confounding factor.

'Relaxed' was included because we considered, based on our experience in earlier pilot studies and on theoretical accounts of the autonomic effects of acupuncture (Mayor 2007; Carpenter et al. 2010; Li et al. 2013; Steffert & Mayor 2014), that relaxation would indicate a positive outcome following acupuncture-based treatments. Although there may be some overlap between 'Comfortable' (including calmness) and 'Relaxation' (Kolcaba 1991), they are not identical, and were therefore both included in NRS-M.

NRS-M does not include a subscale corresponding to $BRUMS_{24}$ Anger because in pilot studies using the POMS-SF, levels of and changes in Anger were not found to be particularly useful measures in the context of electroacupuncture (EA) or transcutaneous electrical acupoint stimulation (TEAS) treatment or teaching (Mayor et al. 2015).

'Anxious' was selected for NRS-M instead of the potentially ambiguous term 'Tension' (physical, mental?), and 'Gloomy' instead of 'Depression' (which could be interpreted as indicating a psychiatric diagnosis). Similarly, the term 'Total Mood Distress' (van Wijk 2011) was preferred to 'Total Mood Disorder' (Lane et al. 2005) for TMD in $BRUMS_{24}$, as being less pathologising.

'Confused' and 'Fatigued' were retained; although, like Anger, 'Confusion' was not found to change greatly following treatment (unpublished pilot data), it was considered as worthy of inclusion in the context of teaching.

Recruitment

Ethics approval was granted under applications for related studies by the Health and Human Sciences Ethics Committee of the University of Hertfordshire, UK (Protocols HEPEC/07/11/93 and HSK/SF/UH/00124). Permission was also received from the course organisers and respondents themselves.

Respondents were recruited during nine pre-arranged EA teaching sessions in the UK. They were not obliged to complete the scales.

Administration

The two measures were administered in paper form in quick succession twice in each teaching session – once early on (NRS-Pre; $BRUMS_{24}$-Pre), when it was judged that most of those who would be attending were present, and once towards the end of the session (NRS-Post; $BRUMS_{24}$-Post), after most participants had received a brief treatment with EA and/or TEAS from a fellow attendee. Order of administration (NRS-M preceding $BRUMS_{24}$, or vice versa) was not consistent over the different sessions. Some attendees declined the treatment because of known contraindications (e.g. pregnancy or a heart condition), and occasionally because of an aversion to something 'non-traditional' like EA or TEAS, or to electricity itself. Some attendees arrived late or left sessions early, and thus were not present to complete both scales. Treatment was supervised but participants were free (within reason) to use their own choice of parameters (frequency, amplitude, mode, pulse and overall stimulation duration) and points.

NRS-M (Appendix A) was printed with subscales positioned on the page in four different orders, so that it was unlikely that they would appear in the same order in both NRS-Pre and NRS-Post for a particular respondent. These two were also distributed and collected separately, so that they could not be seen at the same time. Thus the likelihood of respondents basing their replies to NRS-Post on their earlier replies to NRS-Pre was reduced. $BRUMS_{24}$s were, however, identical in both Pre and Post versions.

Instructions were similar for both scales: 'Place a cross on each line to represent how you feel right now' for NRS-M, and 'circle one of the numbers to the right … to describe how you feel right now' (Appendices A, B). Respondents were asked to create and memorise a personal identification code, and entered this on all the scales they completed. In cases of uncertainty or omission, codes were double-checked against writing style and ink colour.

Analysis

Anonymised scores for each item were entered in an Excel spreadsheet (0-100 for NRS-M, 0-4 for $BRUMS_{24}$). Analysis of the resulting counts for the different scores was carried out in Excel (v10, Microsoft, Redmond, WA) and the Statistical Package for the Social Sciences (SPSS v20, IBM, Armonk, NY). The data was first tested to assess suitability for parametric methods of analysis.

Criterion (concurrent) validity (Objective 1) was assessed using Spearman's *rho* as a measure of association between scales. Convergent and discriminant validity (Objective 2) were compared for the two scales using Spearman's *rho* as a measure of associations within them. Internal consistency (a) and test-retest reliability (b) of NRS-M and $BRUMS_{24}$ (Objective 3) were compared for the two scales using Cronbach's *alpha* and a Friedman's χ^2 (chi-square) test (a) and Pre-Post correlations (*rho*) for the two Scales (b).

Whether or not NRS-M is more sensitive to mood and to changes in mood than $BRUMS_{24}$ (Objective 4, our central objective) was determined using a variety of methods, including Hankins' $delta_G$, comparisons of skewness, counts of low scores and numbers of changes, and Pre-Post comparisons of medians (significance assessed using the Wilcoxon test for paired samples). To determine whether initial state predicts response (Objective 5), Spearman's *rho* was used once again.

Exploratory factor analysis (Objective 6) was conducted using Principal Axis Factoring (as the data did not show multivariate normality), Bartlett's scoring method and direct Oblimin rotation (in case of correlation between factors), with delta set at 0. (Varimax rotation was also investigated, but eventually dropped as a method because direct Oblimin gave higher eigenvalues and, where comparable, often larger loadings.) The Kaiser-Meyer-Olkin (KMO) statistic was calculated and Bartlett's Test of Sphericity conducted, to check that there were patterned relationships between items. Using an eigenvalue cut-off of 1.0, the number of factors that explained cumulative variance and could be retained was then confirmed using scree plots. Hierarchical cluster analysis (Objective 6) was performed by variables on the same data for a range of clusters between 2 and 4. Ward's method was used (with Squared Euclidean distances), and values standardised for the range -1 to +1 (as an attempt to take into account possible negative values for Pre-Post scores). Dendrograms were inspected to aid selection of the most appropriate number of clusters to use from a range of up to 5 solutions.

To investigate common response patterns and styles for NRS-M and $BRUMS_{24}$ (Objective 7), modes were calculated and summed by respondent and by subscale. Response styles (RSs) were calculated as appropriate for our data using standard definitions, counted, and the significance of differences between counts of extreme response style (ERS) responses for the different subscales assessed using the (nonparametric) Kruskal-Wallis test for independent samples. In addition, associations between individual subscale scores and the degree of ERS or midpoint responding (MRS) were considered, using a 0.6 threshold for Spearman's *rho*.

Relationships between ERS and emotional intelligence characteristics at baseline (Objective 8) were explored using *phi* and Goodman and Kruskal's *tau*, with both variables treated as nominal. The Wilcoxon test was again used to compare how long it takes to complete NRS-M and $BRUMS_{24}$.

Results

Respondents

Respondents were recruited from attendees at nine EA training seminars, seven of these being requirements for university-affiliated undergraduate acupuncture training courses (four at the College of Integrated Chinese Medicine, CICM, Reading; two at the Northern College of Acupuncture, NCA, York; one at London South Bank University LSBU), one being a 'continuing professional development' (CPD) course at a summer school held by University of Westminster (UW), and one an independently organised CPD course for established acupuncture practitioners (CPD Group, Guy's Hospital, London). Numbers attending are shown in **Table 2**.

Table 2. Numbers attending and completing scales at the different training seminars.

Course/date	Attend	NRS -Pre	BRUMS -Pre	NRS-Post	BRUMS-Post
CICM 04.08.14	13	13	n/a	12	n/a
CPD 27.11.14	15	15	15	13	13
CICM 24.02.15	26	22	21	26	25
CICM 16.03.15	27	26	27	20	19
NCA 29.03.15	18	18	18	18	18
NCA 24.04.15	16	16	16	14	14
UW 25.07.15	11	8	8	11	11
CICM 17.08.15	19	17	17	16	16
LSBU 22.11.15	13	13	13	12	13
Totals	158	148	135	142	129

Objective 1. *To assess criterion (concurrent) validity of NRS-M by using it in parallel with BRUMS$_{24}$*

NRS-M subscales and BRUMS$_{24}$ items

Some NRS-M subscale names such as 'Anxious', 'Confused' and 'Lively' have exact equivalents in BRUMS$_{24}$ items, but others, such as 'Gloomy' do not. There is no guarantee that 'Gloomy' and DEPR correspond precisely for everybody (particularly if they are not native English speakers), but the strength of association between NRS-M subscales and corresponding BRUMS$_{24}$ items should, at least in theory, be greater than between subscales and items that do not correspond.

Data for most NRS-M and BRUMS$_{24}$ subscales was not found to be normally distributed (see too below, on Skewness). Therefore nonparametric methods of analysis were used throughout. Treating both subscale and item scores as ordinal, Spearman's *rho* was calculated as an index of strength of association ($p<0.001$ in each case). The highest four (absolute) values of *rho* for each NRS-M subscale are shown in **Table 3**.

Table 3. Strongest four associations for each NRS-M subscale

NRS-Pre subscale	**BRUMS-Pre items**
Anxious	Panicky, Anxious, Nervous all >0.6; Worried >0.5
Confused	Confused >0.7; Mixed up >0.5; Muddled & Uncertain >0.4
Fatigued	Exhausted > 0.7; Worn out, Sleepy & Tired >0.6
Gloomy	Downhearted > 0.5; Depressed, Unhappy, Miserable & *Bad tempered* > 0.4
Lively	Lively & Energetic > 0.6; Active > 0.5; Alert & *All 4 FAT items** > 0.4
NRS-Post subscale	**BRUMS-Post items**
Anxious	Anxious > 0.6; Panicky, Worried, Nervous & *Mixed up* > 0.5
Confused	Confused & Mixed up > 0.6; Uncertain > 0.5; Muddled & *Panicky* > 0.4
Fatigued	Exhausted & Tired > 0.6; Sleepy & Worn out > 0.5
Gloomy	Depressed > 0.6; Downhearted & *Mixed up* > 0.5; Miserable, Unhappy, *Nervous & Uncertain* > 0.4
Lively	Lively & Energetic > 0.7; Active > 0.6; Alert & *Exhausted, Tired** > 0.4

* Negative associations.

Values of *rho* thus confirm that both NRS-Pre and NRS-Post subscales do correspond to the items for the corresponding BRUMS$_{24}$ subscales, although for some BRUMS$_{24}$ items *rho* is <0.5. However, there is also a negative association between NRS-M 'Lively' and BRUMS$_{24}$ FAT items (see further on this below). In addition, 'Gloomy' is associated with the BRUMS$_{24}$ Bad tempered item in the Pre scales, and with both BRUMS$_{24}$ Nervous and Mixed up/Uncertain items in the Post scales (with a further association between 'Confused' and BRUMS$_{24}$ Panicky).

The correlations between 'Anxious' and BRUMS$_{24}$ TENSion items are close to that reported by other authors using a VAS for anxiety and POMS TENSion (r=0.51; Cella & Perry 1986).

Table 4 shows correspondences for the Pre-Post differences (Post minus Pre values).

Table 4. Strongest associations between each NRS-M Pre-Post scale and $BRUMS_{24}$ item score differences (omitted for *rho* < 0.4).

NRS Pre-Post differences (subscales)	**$BRUMS_{24}$ Pre-Post differences (items)**
Anxious	Nervous >0.5; Anxious, Worried, *Downhearted* & *All 4 FAT items* >0.4 [Panicky > 0.3]
Confused	Confused >0.5; Mixed up >0.4 [Muddled & Uncertain > 0.3]
Fatigued	Exhausted, Tired, Worn out >0.5; Sleepy & *Energetic** > 0.4
Gloomy	*Worn out* > 0.4 [Depressed & Downhearted > 0.3; Miserable > 0.2; Unhappy < 0.1]
Lively	Lively & Energetic > 0.5; *Tired** > 0.4 [Active & Alert > 0.3]

* Negative associations.

With the differences between Pre and Post scores, the associations between NRS-M and $BRUMS_{24}$ become less convincing, particularly for 'Gloomy'.

NRS-M subscales and $BRUMS_{24}$ subscales

Table 5 shows correspondences between NRS-M and BRUMS24 subscales, and between 'Overall mood' or 'Total mood' ('Anxious' + 'Confused' + 'Fatigued' + 'Gloomy' – 'Lively') and TMD or TMD-A (TMD but omitting the Anger subscale score).

Table 5. Associations for each NRS-M subscale and for 'Overall mood.

NRS-Pre	**BRUMS-Pre (same subscale)**	**BRUMS-Pre (other subscales)**
Anxious	TENSion > 0.7	Confusion & Depression > 0.5
Confused	CONFusion > 0.6	
Fatigued	FATigue > 0.7	VIGour* > 0.5; CONFusion & DEPRession > 0.4
Gloomy	DEPRession > 0.5	ANGer > 0.4
Lively	VIGour > 0.6	FATigue* > 0.4
Overall mood	TMD*, TMD-A* > 0.6	VIGour > 0.6; DEPRession* > 0.5; ANGer*, CONFusion*, FATigue* > 0.4
Total mood	TMD & TMD-A > 0.8	CONFusion, DEPRession, FATigue > 0.6; TENSion > 0.5; ANGer > 0.4
NRS-Post	**BRUMS-Post (same subscale)**	**BRUMS-Post (other subscales)**
Anxious	TENSion > 0.6	Confusion & Depression > 0.5
Confused	CONFusion > 0.6	TENSion > 0.4
Fatigued	FATigue > 0.6	
Gloomy	DEPRession > 0.6	CONFusion > 0.5
Lively	VIGour > 0.7	FATigue* > 0.4
Overall mood	TMD*, TMD-A* > 0.7	CONFusion > 0.6; DEPRession*, TENSion* > 0.5; FATigue* > 0.4
Total mood	TMD & TMD-A > 0.7	CONFusion > 0.6; DEPRession & TENSion > 0.5; FATigue > 0.4

* Negative associations.

Values of *rho* confirm that both NRS-Pre and NRS-Post subscales correlate well with the corresponding $BRUMS_{24}$ subscales, indeed better than NRS-M subscales with $BRUMS_{24}$ items. Both Overall and Total mood correlate well with TMD and TMD-A.

Table 6 shows correspondences for the Pre-Post NRS-M and $BRUMS_{24}$ subscale differences.

Table 6. Strongest associations between each NRS-M pre-Post subscale and $BRUMS_{24}$ subscale differences (omitted for *rho* < 0.4).

NRS-M Pre-Post	**$BRUMS_{24}$ Pre-Post (same subscale)**	**$BRUMS_{24}$ Pre-Post (other subscales)**
Anxious	TENSion > 0.6	FATigue > 0.5; CONFusion & DEPRession > 0.4
Confused	CONFusion > 0.6	FATigue > 0.4
Fatigued	FATigue > 0.7	CONFusion, DEPRession & TENSion > 0.4
Gloomy	DEPRession > 0.4	CONFusion & FATigue > 0.4
Lively	VIGour > 0.6	
Overall mood		VIGour > 0.5
Total mood	TMD & TMD-A > 0.7	FATigue > 0.6; CONFusion > 0.5; DEPRession & TENSion > 0.4

Results are least convincing here for 'Gloomy', as in **Tables 3** (for BRUMS-Post items) and **4** (for Pre-Post item differences). Note that of the four $BRUMS_{24}$ DEPRession items in **Tables 3** and **4**, *rho* for 'Gloomy' is highest for Downhearted (BRUMS-Pre items) and for Depressed and Downhearted (BRUMS-Post items and Pre-Post item differences).

These findings can be compared with those of Terry et al. (2003), who considered that the correlations they found between $BRUMS_{24}$ and the original POMS subscales indicated criterion validity of the former. Their results are shown in **Table 7 (a)** (using Pearson's r rather than Spearman's *rho*, although non-normality was found for some variables in their study), with NRS-M/$BRUMS_{24}$ correlations in **Table 7 (b)**.

Table 7. Criterion validity of $BRUMS_{24}$ against the original POMS (a), and of NRS-M against $BRUMS_{24}$ subscales (b). The three values of *rho* are for the Pre, Post and Pre-Post subscale comparisons.

$BRUMS_{24}$ / POMS	**(a) Pearson's r**	**NRS / $BRUMS_{24}$**	**(b) Spearman's *rho***
ANGer	>0.8	n/a	n/a
CONFusion	>0.7	Confused/CONF	> 0.6; > 0.6; > 0.6
DEPRession	>0.8	Gloomy/DEPR	>0.5; > 0.6; > 0.4
FATigue	=0.9	Fatigued/FAT	> 0.7; > 0.6; > 0.7
TENSion	>0.7	Anxious/TENS	> 0.7; > 0.6; > 0.6
VIGour	>0.6	Lively/VIG	> 0.6; > 0.7; > 0.6

Given that NRS-M uses a very different format to both the POMS and $BRUMS_{24}$, these results are encouraging.

Data is currently being collected to enable assessment of the external (criterion) validity of NRS-M in relation to measures such as HRV.

Summary 1 – *Concurrent validity of NRS-M*

- Nonparametric correlation confirms that both NRS-Pre and NRS-Post subscales correspond to the items for the corresponding $BRUMS_{24}$ subscales, although for some $BRUMS_{24}$ items *rho* is <0.5.
- Correlations of NRS-Pre and NRS-Post subscales with the corresponding $BRUMS_{24}$ subscales are good (>0.6 or above), indeed better than NRS-M subscales with $BRUMS_{24}$ items.
- Data is currently being collected to enable assessment of the external (criterion) validity of NRS-M in relation to measures such as HRV.

Objective 2. *To explore convergent and discriminant validity of NRS-M compared with* $BRUMS_{24}$

How useful NRS-M proves in practice will depend on its convergent and discriminant validity, namely how well it indicates correspondence/convergence between theoretically similar constructs, or discriminates between constructs that are theoretically different (DeVon et al. 2007). Within-scale significant correlations (*rho* > 0.3) are shown in **Tables 8-10**, with the strongest correlations in red.

Table 8. Within scale correlations (*rho* > 0.3) for NRS-Pre.

NRS-Pre	Anx	Comf	Conf	Fatig	Gloom	Lively	Relax	Overall	Total
Anx		-0.3	0.4		0.4		-0.3	-0.4	0.7
Comf						0.4	0.7	0.6	-0.5
Conf								-0.3	0.6
Fatig					0.3	-0.4		-0.4	0.7
Gloom							-0.3	-0.5	0.6
Lively							0.3	0.5	-0.5
Relax								0.6	-0.3
Overall									-0.6
Total									

Table 9. Within scale correlations (*rho* > 0.3) for NRS-Post.

NRS-Post	Anx	Comf	Conf	Fatig	Gloom	Lively	Relax	Overall	Total
Anx		-0.3	0.5		0.5		-0.4	-0.4	0.6
Comf						0.3	0.8	0.7	-0.7
Conf					0.5		-0.3	-0.3	0.6
Fatig					0.3	-0.3			0.5
Gloom								-0.4	0.6
Lively							0.3	0.4	-0.5
Relax								0.7	-0.7
Overall									-0.7
Total									

Table 10. Within scale correlations (*rho* > 0.3) for NRS Pre-Post.

NRS Pre-Post	Anx	Comf	Conf	Fatig	Gloom	Lively	Relax	Overall	Total
Anx			0.4	0.4	0.5				0.6
Comf						0.5	0.7	0.5	
Conf				0.3					0.5
Fatig					0.4				0.7
Gloom									0.6
Lively							0.4	0.6	-0.3
Relax								0.6	
Overall									
Total									

In all three of the **Tables**, the highest correlation was between 'Comfortable' and 'Relaxed' ('Overall mood' also showed strong correlations with 'Comfortable'/'Relaxed'). Other correlations greater than 0.5 were found for NRS-Post between 'Anxious', 'Confused' and 'Gloomy'.

For both NRS-Pre and NRS-Post, low negative correlations were found between 'Anxious' and 'Comfortable'/'Relaxed', 'Fatigued' and 'Lively', with additional negative correlations between 'Gloomy' and 'Relaxed' (NRS-Pre) and 'Confused' and 'Relaxed' (NRS-Post).

Comparable results for $BRUMS_{24}$ are shown in **Tables 11-13**.

Table 11. Within scale correlations (*rho* > 0.3) for BRUMS-Pre.

BRUMS-Pre	ANG	CONF	DEPR	FAT	TENS	VIG	TMD	TMD-A
ANG		0.8	0.6	0.3	0.4		0.6	0.5
CONF			0.5	0.5	0.5		0.7	0.7
DEPR				0.4	0.5		0.7	0.7
FAT					0.4		0.7	0.8
TENS							0.6	0.6
VIG							-0.4	-0.4
TMD								0.9
TMD-A								

Table 12. Within scale correlations (*rho* > 0.3) for BRUMS-Post.

BRUMS-Pre	ANG	CONF	DEPR	FAT	TENS	VIG	TMD	TMD-A
ANG		0.3	0.5		0.4		0.4	0.4
CONF			0.5	0.3	0.6		0.6	0.6
DEPR				0.3	0.6		0.6	0.6
FAT					0.3		0.7	0.7
TENS							0.6	0.6
VIG							-0.4	-0.4
TMD								0.9
TMD-A								

Table 13. Within scale correlations (*rho* > 0.3) for $BRUMS_{24}$ Pre-Post.

BRUMS Pre-Post	ANG	CONF	DEPR	FAT	TENS	VIG	TMD	TMD-A
ANG		0.3	0.6	0.3	0.4		0.5	0.4
CONF			0.3	0.4	0.4		0.6	0.6
DEPR				0.4	0.3		0.6	0.6
FAT					0.4		0.8	0.8
TENS							0.5	0.5
VIG							-0.3	-0.3
TMD								0.9
TMD-A								

For $BRUMS_{24}$ high *rho* are found between ANGer and DEPRession (0.6, 0.5, 0.6), but also between CONFusion and DEPRession (0.5, 0.5, [0.3]), between TENSion and CONFusion (0.5, 0.6, [0.4]), and between DEPRession and CONFusion (0.5, 0.5, [0.3]). No correlations with VIGour were strong, and the consistently strongest correlations with both TMD and TMD-A were with FATigue.

Table 14 shows the means (SDs) of *rho* >0.3 for those correlations between equivalent subscales that occur in both NRS-M and $BRUMS_{24}$ (i.e excluding 'Comfortable', 'Relaxed' and ANGer) and between these subscales and equivalents of 'Overall mood'/TMD.

Table 14. Means (SDs) of *rho* >0.3 for correlations between equivalent subscales in NRS-M and $BRUMS_{24}$.

Scale	Between subscales	Between Overall/TMD & subscales	Between Total/TMD-A & subscales
NRS-Pre	0.38 (0.05)	0.42 (0.08)	0.62 (0.08)
BRUMS-Pre	0.47 (0.05)	0.62 (0.13)	0.64 (0.15)
NRS-Post	0.42 (0.11)	0.38 (0.05)	0.56 (0.05)
BRUMS-Post	0.43 (0.15)	0.58 (0.11)	0.58 (0.11)
NRS Pre-Post	0.40 (0.07)	0.60 (0.00)	0.54 (0.15)
BRUMS Pre-Post	0.37 (0.05)	0.56 (0.18)	0.56 (0.18)

Thus, crudely, $BRUMS_{24}$ shows greater mean correlation between subscales, and between subscales and total scores, than does NRS-M (except for two out of three of the Pre-Post means). If the constructs underlying the two scales are more or less equivalent (except for those excluded here), and are indeed different from each other, this suggests that NRS-M shows greater discriminant validity than $BRUMS_{24}$.

In NRS-M, the highest correlations for positive valence 'moods' were between 'Comfortable' and 'Relaxed', and for negative valence moods between 'Anxious', 'Confused' and 'Gloomy'. For $BRUMS_{24}$, strong correlations were similarly found among ANGer, CONFusion, DEPRession and TENSion. NRS-M would appear to show good convergent validity for the possibly related terms 'Comfortable' and Relaxed' (see above), and $BRUMS_{24}$ for ANGer and DEPRession, which are related in several theoretical models of depression (Luutonen 2007; Maciocia 2009). Both scales showed some association between 'Anxious'/TENSion, 'Confused'/CONFusion and 'Gloomy'/DEPRession – although not to the extent that they could be considered as a single mood entity, or even as overlapping.

Small negative correlations were shown between 'Anxious' and 'Relaxed' (as also between TENSion and 'Relaxed'), as would be expected.

Summary 2 – *Convergent and discriminant validity of NRS-M*

- Nonparametric correlation confirms convergence between 'Comfortable' and 'Relaxed'.
- $BRUMS_{24}$ shows greater mean correlation between subscales than NRS-M.
- This suggests that NRS-M shows greater discriminant validity than $BRUMS_{24}$.

Objective 3. *To compare internal consistency and test-retest reliability of NRS-M and $BRUMS_{24}$*

Cronbach's *alpha*

Given that the different subscales of NRS-M were selected precisely because they were considered to measure different moods, it is not really meaningful to calculate Cronbach's *alpha* for the scale as a whole. This is confirmed by consistently low values of *alpha* for both NRS-Pre and NRS-Post, but significant results for Friedman's χ^2 (chi-square) tests ($p<0.001$). *Alpha* was greatest when calculated for the positive subscales only (with or without 'Overall mood'), least for all Subscales (including 'Overall mood') taken together. For $BRUMS_{24}$, Cronbach's *alpha* was considerably higher (e.g. 0.694 for BRUMS-Pre vs 0.234 for NRS-Pre, and 0.634 for BRUMS-Post vs 0.404 for NRS-Post – including TMD/'Overall mood'). Internal consistency for the individual $BRUMS_{24}$ subscales was of course far higher, being greatest for FATigue in both BRUMS-Pre and BRUMS-Post (**Table 15**).

Table 15. Cronbach's *alpha* for the individual $BRUMS_{24}$ subscales.

	ANG	CONF	DEP	FAT	TENS	VIG
BRUMS-Pre	0.856	0.819	0.881	0.927	0.900	0.875
N	131	132	133	134	135	135
BRUMS-Post	0.809	0.874	0.895	0.928	0.876	0.893
N	127	125	126	126	128	125

However, it should be noted that 'reliable measures may fail to discriminate adequately, and developing measures solely to maximise internal reliability may be counterproductive' (Hankins 2008). Test discrimination $delta_G$ (see below) is good for NRS-M, and this may compensate for low values of Cronbach's *alpha*, whereas for $BRUMS_{24}$, although internal consistency is good, its test discrimination is less so.

Test-retest reliability

If NRS-M and $BRUMS_{24}$ scales are sensitive to changes in mood, their test-retest reliability should not be high when they are administered before and after an intervention.

Pre-Post correlations (*rho*) for the two Scales are shown in **Table 16**.

Table 16. Pre-Post correlations (*rho*) for NRS-M and $BRUMS_{24}$.

	Anx/TENS	Comf/-	Conf	Fat	Gl/DEPR	Lively/V	Rel/-	-/ANG	Ov/TMD
NRS-M	0.470	0.527	0.435	0.336	0.404	**0.238**	0.470	n/a	0.575
$BRUMS_{24}$	0.455	n/a	**0.240**	0.344	0.487	0.363	n/a	0.306	**0.411**

Pre-Post correlations are <0.5 for all except NRS-M 'Comfortable' and 'Overall mood'. Greatest differences between the two scales are for 'Confused'/CONFusion and 'Lively'/VIGour. For the former subscale, $BRUMS_{24}$ shows lower reliability (greater sensitivity?), and for the latter NRS-M shows lower reliability (greater sensitivity?).

In a study on POMS and other neuropsychological measures, the lowest test-retest reliability over several weeks was found for POMS FATigue at r=0.39 (Salinsky et al. 2001). This compares surprisingly well with the $BRUMS_{24}$ result here. For POMS TENSion, test-retest reliability was r=0.70 in another study (Rossi & Pourtois 2011) when no treatment was received between applications, but r=0.51 when 'external modifications' (e.g. treatment) intervened before the retest phase – again, this is not too different from the $BRUMS_{24}$ result here. Test-retest reliability for a single VAS for anxiety in healthy populations has been found to be as low as 0.30 or >0.50 in different studies (Rossi & Pourtois 2011).

Data is currently being collected to enable assessment of stability (test-retest) reliability of NRS-M when administered repeatedly before different interventions, a week or more apart.

Summary 3 – *Internal consistency and test-retest reliability of NRS-M*

- Cronbach's *alpha* was consistently low for both NRS-Pre and NRS-Post, greatest when calculated for positive subscales only (with or without 'Overall mood').
- Cronbach's *alpha* was considerably – and predictably – higher for $BRUMS_{24}$.
- However, 'developing measures solely to maximise internal reliability may be counterproductive' (Hankins 2008), and test discrimination $delta_G$ may compensate for low *alpha*.
- In this context, test-retest reliability was low (except for 'Comfortable'), possibly indicating greater sensitivity than $BRUMS_{24}$ to changes in some moods.
- Data is currently being collected to enable assessment of stability (test-retest) reliability of NRS-M when administered repeatedly at intervals of a week or more.

Objective 4. *To determine whether NRS-M is more sensitive to mood and to changes in mood than BRUMS$_{24}$ using a variety of methods*

Sensitivity to mood

Method 1. *Delta$_G$*

A useful initial indicator of a scale's sensitivity is its ability to discriminate between individuals *without* reference to an external criterion (its test discrimination) (Hankins 2008). One early test discrimination measure was Ferguson's *delta* for dichotomous data (Ferguson 1949). Hankins (2007) generalised this as *delta$_G$* to enable its use with polytomous data, as in the Likert-based BRUMS$_{24}$ and NRS-M examined here, although his approach has not been without its critics (Norman 2008; Wyrwich 2008). In this context, 'as *delta* varies as a function of the distribution then mood might reasonably be expected to affect it' (Hankins. Personal communication, 17 Feb 2016). In other words, a measure of moods which discriminates better between the moods of individuals is probably going to be more sensitive to the moods themselves.

Using *delta$_G$* on both NRS-M and BRUMS$_{24}$ provided the results shown in **Table 17**.

Table 17. Test discrimination *delta$_G$* for NRS-M and BRUMS$_{24}$.
Shown in **bold** are those results with high *delta$_G$* and greatest Pre-Post difference,
and in red where *delta$_G$* was higher in Post than Pre measures.

NRS subscale	NRS-Pre	NRS-Post	BRUMS subscale	BRUMS-Pre	BRUMS-Post
Anxious	0.9373	0.8247	TENSion	0.6756	0.6435
Comfortable	0.9607	0.9502	[no equiv't]		
Confused	0.9419	0.8868	CONFusion	0.7179	0.6748
Fatigued	0.9804	0.9620	FATigue	0.8363	0.9568
Gloomy	0.9111	0.7401	DEPRession	0.5297	0.3788
Lively	**0.9657**	**0.9650**	VIGour	**0.7832**	**0.9342**
Relaxed	0.9548	0.9451	[no equiv't]		
[no equiv't]			ANGer	0.4605	0.2961
Overall mood	0.9306	0.9136	[no equiv't]		
TOTAL	0.9044	0.9398	TMD	0.7316	0.7796

Delta$_G$ is consistently greater for NRS-M than for BRUMS$_{24}$, both Pre and Post, suggesting that the former has greater sensitivity. The mean percentage differences in *delta$_G$* for comparable NRS-M and BRUMS$_{24}$ subscales are 24.4% (Pre) and 19.2% (Post).

Delta$_G$ is also consistently less for NRS-Post than NRS-Pre (except for TOTAL), and similarly for most BRUMS$_{24}$ subscales except FATigue and VIGour (and also for TMD). This suggests that, for whatever reason, the scale's ability to discriminate between individuals – and also its sensitivity – may be greater earlier than later in a teaching session.

It could be argued that an 11-point scale (which is basically what NRS-M subscales are, if values between multiples of ten are recalculated to the nearest multiple) is more likely to detect differences than a measure such as BRUMS$_{24}$, constructed on 5-point Likert scales. On the other

hand, as there are four items for every $BRUMS_{24}$ subscale, it could also be argued that each subscale has potentially 4 x 5, or 20, data points, so would be more likely to detect differences. Whatever the interpretation, clearly more precise ways of determining sensitivity should be used in addition to $delta_G$, which is only an indicator of *likely*, rather than actual, sensitivity.

Method 2. Skewness

The asymmetric subscales used in both NRS-M and $BRUMS_{24}$ are very likely to result in skewed distributions (Lorr 1989).

Skewness for the various subscales was calculated (divided by its standard error, SE), and highlights some consistent patterns in how they were scored (**Fig 1**).

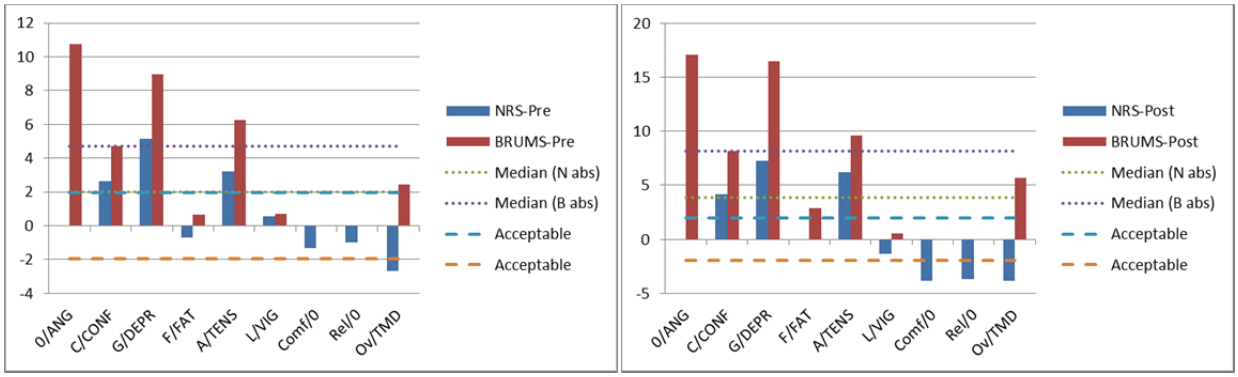


Figure 1. Skewness/SE of NRS-M and $BRUMS_{24}$, consistently positive and almost always numerically greater for $BRUMS_{24}$.

Of NRS scales, NRS-Post 'Comfortable', 'Relaxed' and 'Overall mood' were all negatively skewed (to the right), whereas for NRS-Pre this was only so for 'Overall mood'. In contrast, 'Anxious', 'Confused' and 'Gloomy' were all positively skewed (to the left), more so for NRS-Post than NRS-Pre subscales. The other subscales exhibited acceptable skewness, with absolute skewness divided by its standard error being <1.96 (Abu-Bader 2010).

Of $BRUMS_{24}$ subscales, only VIGour and BRUMS-Pre FATigue showed acceptable skewness. All the others were positively skewed (to the left), some markedly so.

Furthermore, (absolute) values of skewness were greater for $BRUMS_{24}$ than for NRS-M.

Thus the results for $BRUMS_{24}$ exhibit greater bias and are probably less dependable than those for NRS-M.

Skewness/SE is also consistently greater for the Post than the Pre subscales – except for $BRUMS_{24}$ VIGour and NRS-M Fatigued. This is in line with the results for $delta_G$.

Method 3. Counts of 'low scores'

Counts were made of 'low scores' for each subscale (defined as <10% of the maximum possible score of the subscale, i.e. <10 for NRS-M and <1.6 for $BRUMS_{24}$, or 8 for TMD). To be able to compare the counts 'n' for the two measures, they were normalised by dividing by the number of respondents for that subscale (*N*). Results are shown in **Table 18**.

Table 18. Counts of 'low scores' for NRS-M and $BRUMS_{24}$ (n), with normalised 'low scores' (n/*N*).

NRS-Pre			**$BRUMS_{24}$-Pre**		
NRS item (*N*=148)	n	n/*N*	$BRUMS_{24}$ subscale items (*N*=135)	n	n/*N*
Anxious	41	0.277	Tension	52	0.380
Comfortable	1	0.007	[no equivalent]	n/a	n/a
Confused	38	0.257	Confusion	48	0.350
Fatigued	9	0.061	Fatigue	16	0.117
Gloomy	48	0.324	Depression	82	0.599
Lively	4	0.027	Vigour	9	0.066
Relaxed	0	0	[no equivalent]	n/a	n/a
[no equivalent]	n/a	n/a	Anger	93	0.679
Overall mood	0	0	TMD	57	0.416
Sum	141	0.953	Sum (not including TMD)	300	2.190
NRS-Post			**$BRUMS_{24}$-Post**		
NRS item (*N*=142)	n	n/*N*	$BRUMS_{24}$ subscale items (*N*=129)	n	n/*N*
Anxious	62	0.437	Tension	72	0.558
Comfortable	4	0.028	[no equivalent]	n/a	n/a
Confused	48	0.338	Confusion	73	0.566
Fatigued	9	0.063	Fatigue	18	0.140
Gloomy	74	0.521	Depression	102	0.791
Lively	3	0.021	Vigour	10	0.078
Relaxed	3	0.021	[no equivalent]	n/a	n/a
[no equivalent]	n/a	n/a	Anger	111	0.860
Overall mood	0	0	TMD	81	0.628
Sum	203	1.430	Sum (not including TMD)	386	2.992

Comparing numbers of low scores per individual (n/*N*), the Likert-based $BRUMS_{24}$ is **less sensitive to all moods** than NRS-M, with the ratio of n/*N* for $BRUMS_{24}$ to that for NRS varying between 1.27 (TENS/'Anxious', Post) to 2.44 (VIG/'Lively' Pre). Where similar moods are being compared, this difference in sensitivity is significant only for Pre 'Gloomy'/DEPR ($p<0.001$) and Post 'Confused'/CONF ($p=0.004$) and 'Gloomy'/DEPR ($p=0.005$) – using the Binomial test with test proportion 0.48 or 135/(135+148) for Pre scales, and 0.48 or 129/(129+142) for Post scales.

For both scales, the ratio of n/*N* for scale-Pre to that for scale-Post is consistently <1, except for 'Lively' (the ratio for VIG, although <1, is the greatest of all the $BRUMS_{24}$ ratios at 0.846). As with Methods 1 and 2 above, this possibly indicates less sensitivity to mood towards the end of the teaching session than at the beginning of the session.

Instead of n/*N*, the percentage of low scores for all subscales taken together can be calculated from **Table 18** as:

Percentage of low scores for each scale = Sum/(*N* x number of subscales)%.

Results again indicate that $BRUMS_{24}$ is less sensitive than NRS-M (**Table 19**).

Table 19. Percentage of low scores for all subscales taken together, with significance of differences in low scores between the two measures, using the Binomial test.

NRS-Pre	BRUMS-Pre	significance
11.9	31.7	p=0.001
NRS-Post	BRUMS-Post	significance
17.9	42.7	p<0.001

Sensitivity to changes in mood

Method 1. *$Delta_G$*

$Delta_G$ may be calculated for Pre-Post score differences as well as for the scores themselves, and again indicates greater test discrimination for NRS-M than $BRUMS_{24}$. Using absolute rather than signed differences, *$delta_G$* for the Pre-Post score differences are shown in **Table 20**. (Results for the signed differences also indicate greater test discrimination for NRS-M, except for 'Confused'/CONF, where *$delta_G$* is slightly (1.3%) greater for $BRUMS_{24}$.

Table 20. Test discrimination *$delta_G$* for NRS-M and $BRUMS_{24}$ Pre-Post score differences.

NRS subscale	NRS Pre-Post	BRUMS subscale	BRUMS Pre-Post
Anxious	0.8614	TENSion	0.8173
Comfortable	0.8431	[no equivalent]	
Confused	0.8774	CONFusion	0.8461
Fatigued	**0.9167**	FATigue	0.8973
Gloomy	0.8799	DEPRession	0.7007
Lively	0.9066	VIGour	0.8610
Relaxed	0.8311	[no equivalent]	
[no equivalent]		ANGer	0.6946
Overall mood	0.7817	[no equivalent]	
TOTAL	0.9037	TMD	0.8307

Method 2. Skewness

Although skewness for Pre and Post subscales appears to favour NRS-M, for the Pre-Post score differences it is surprisingly good for both NRS-M and $BRUMS_{24}$ (**Fig 2**). Median skewness/SE is marginally better at 1.61 for NRS-M than at 1.91 for $BRUMS_{24}$ (both these being acceptable).

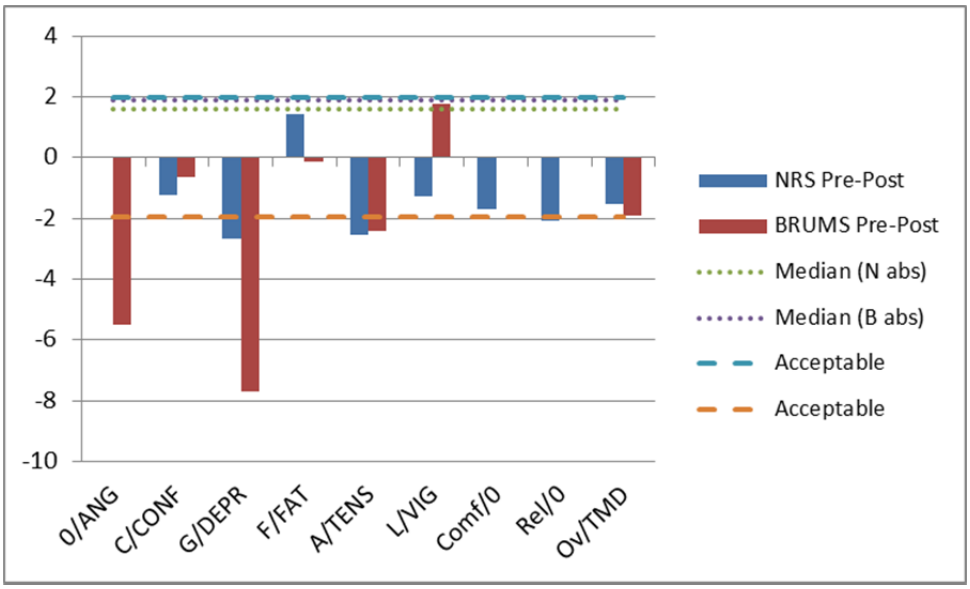


Figure 2. Skewness of NRS-M and $BRUMS_{24}$ Pre-Post score differences, mostly negative for both scales, and very different only for 'Gloomy'/DEPRessed and 'Lively'/VIGour.

Method 3. Counts of 'low scores'

Small *differences* between Pre and Post mood scores would indicate that the scale is not sensitive to changes in mood in this context. Results from using Method 3 as above, but for the 'Post minus Pre' difference scores, are shown in **Table 21**.

Table 21. Counts of 'low difference scores' (Post minus Pre) for NRS-M and $BRUMS_{24}$ (n), with normalised 'low scores' (n/*N*).

NRS			**$BRUMS_{24}$**		
NRS item (*N*=132)	n	n/*N*	$BRUMS_{24}$ subscale (*N*=121)	n	n/*N*
Anxious	45	0.341	Tension	64	0.529
Comfortable	29	0.220	[no equivalent]		
Confused	43	0.159	Confusion	57	0.471
Fatigued	19	0.144	Fatigue	38	0.314
Gloomy	49	0.371	Depression	83	0.686
Lively	21	0.159	Vigour	52	0.430
Relaxed	32	0.242	[no equivalent]		
[no equivalent]			Anger	86	0.711
Overall mood	43	0.326	TMD	56	0.463
Sum	281	1.962	Sum (not including TMD)	380	3.141

Here there are fewer 'low' NRS than $BRUMS_{24}$ differences for all corresponding moods in the two scores. Using a test proportion of 132/(132+121), or 0.52, the Binomial test gives the following as significant:

Anxious/TENS	p=0.016	Overall/TMD	n.s.
Confused/CONF	p=0.044		
Fatigued/FAT	p=0.003		
Gloomy/DEPR	p<0.001		
Lively/VIG	p<0.001		

Thus, NRS is more sensitive to change than $BRUMS_{24}$ for all comparable subscales (although not significantly so for 'Overall mood'/TMD).

Percentages of low difference scores for each scale again indicate that $BRUMS_{24}$ is less sensitive than NRS-M (**Table 22**).

Table 22. Percentage of low difference scores for all subscales taken together, together with significance of the differences in low scores between the two measures, using the Binomial test.

NRS Pre-Post	BRUMS Pre-Post	significance of difference
26.6	44.9	p=0.009

Thus NRS Pre-Post difference is sensitive to around 73% of mood changes, as compared with BRUMS Pre-Post difference, sensitive to around 55%.

Method 4. Numbers of changes

Rather than concentrating on how many low difference scores there are, the numbers of changes (increases and decreases) can be considered. If changes were random, approximately equal numbers of increases and decreases would be expected. Binomial tests were conducted to determine whether the ratio of increases to decreases was in fact significantly different from that expected by chance. Results are shown in **Table 23** for the different subscales of NRS-M and $BRUMS_{24}$.

Table 23. Numbers of changes (increases, decreases and no changes) for NRS-M and $BRUMS_{24}$ subscales, together with significance of the ratios of increases to decreases (using the Binomial test). Arrows indicate predominant direction of change. Column maxima are in **bold**.

NRS Pre-Post	inc	dec	0	Binom p	ratio	BRUMS Pre-Post	inc	dec	0	Binom p	ratio
Anxious	30	84	44	<0.001↓	0.36	Tension	28	73	44	<0.001↓	0.38
Comfortable	68	64	26	ns ↑	1.06	[no equiv't]					
Confused	40	77	41	0.001↓	0.52	Confusion	33	**84**	28	<0.001↓	0.39
Fatigued	48	**95**	15	<0.001↓	0.51	Fatigue	47	**84**	14	0.002↓	0.56
Gloomy	31	80	47	<0.001↓	0.39	Depression	12	58	75	<0.001↓	0.21
Lively	76	68	14	ns ↑	1.12	Vigour	**58**	65	22	ns↓	0.89
Relaxed	**78**	55	25	ns ↑	1.42	[no equiv't]					
[no equiv't]						Anger	17	53	75	<0.001↓	0.32
Overall mood	73	56	29	ns ↑	1.30	TMD	34	**93**	31	<0.001↓	0.37
Sum	444	579	241	<0.001↓	0.77	Sum (not incl TMD)	195	417	258	<0.001↓	0.47

There are more increases and decreases for the 7-subscale NRS-M than for the 6-subscale $BRUMS_{24}$ subscales, as well as total numbers of changes. If only the five comparable subscales are considered ('Anxious'/TENS, 'Confused'/CONF, 'Gloomy'/DEPR, 'Fatigued'/FAT and 'Lively'/VIG), and a correction introduced to allow for the differences in sample size (132 for NRS Pre-Post, 121 for

BRUMS Pre-post), then the results are as shown in **Table 24**. These allow a more accurate comparison of results for the two scales.

Table 24. Numbers of changes (increases, decreases and no changes) for the five comparable NRS-M and $BRUMS_{24}$ subscales, together with significance of the ratios of increases to decreases (using the Binomial test). Arrows indicate predominant direction of change.

NRS Pre-Post	inc	dec	0	Binom p	ratio	BRUMS Pre-Post	inc	dec	0	Binom p	ratio
5 subscales	225	404	161	<0.001↓	0.56	5 subscales	194	397	200	<0.001↓	0.49

There is now no significant difference between numbers of increases or decreases for the two scales, but there are significantly more 'no change' results for $BRUMS_{24}$ than for NRS-M (p=0.045).

Method 5. Comparing median values for Pre and Post scales

Table 25 shows the median values for each Pre and Post subscale where these were both completed by respondents, together with normalised percentage changes.

Table 25. Median values for each Pre and Post subscale where these were both completed by respondents, together with normalised percentage changes, namely (Post-Pre)/Pre%.

NRS subscale	NRS-Pre	NRS-Post	%Diff	BRUMS subscale	BRUMS-Pre	BRUMS-Post	%Diff
Anxious (*N*=132)	20	10	-50.0	TENSion (*N*=120)	3	1	-66.7
Comfortable (*N*=132)	70	70	0.0	[no equiv't]			
Confused (*N*=131)	30	10	-66.7	CONFusion (*N*=120)	3	1	-66.7
Fatigued (*N*=132)	53	50	-5.7	FATigue (*N*=120)	7	6	-14.3
Gloomy (*N*=131)	13	5	-61.5	DEPRession (*N*=120)	1	0	-100.0
Lively (*N*=132)	50	50	0	VIGour (*N*=120)	7	7	0.0
Relaxed (*N*=132)	60	70	+16.7	[no equiv't]			
[no equiv't]				ANGer (*N*=120)	0	0	0.0
Overall mood (*N*=131)	70	70	0.0	[no equiv't]			
Median all	50	50	0.0	TMD	10	3	-70.0

In contrast to the previous methods, the normalised percentage median changes here suggest that – apart from 'Confused'/CONF and 'Lively'/VIG, $BRUMS_{24}$ is more sensitive to changes in mood than NRS-M.

Method 6. Testing for significance of the differences found using Method 5

Using the Wilcoxon test for paired samples, 2-tailed asymptotic significance of the Pre-Post differences for the two scales are shown in **Table 26**, together with *Z* scores and effect size *r* (DeBruine n.d.).

Table 26. Results of the Wilcoxon test for paired (Pre/Post) samples, , together with *Z* scores and effect size *r*.

NRS subscale	*Z*	p value	Effect size *r*	BRUMS subscale	*Z*	p value	Effect size *r*
Anxious (*N*=132)	4.30	<0.001	0.37	TENSion (*N*=120)	4.28	<0.001	0.39
Comfortable (*N*=132)	0.99	ns	0.09	[no equiv't]			
Confused (*N*=131)	3.74	<0.001	0.33	CONFusion (*N*=120)	4.57	<0.001	0.42
Fatigued (*N*=132)	3.18	0.001	0.28	FATigue (*N*=120)	2.64	0.008	0.24
Gloomy (*N*=131)	3.88	<0.001	0.34	DEPRession (*N*=120)	4.85	<0.001	0.44
Lively (*N*=132)	1.13	ns	0.10	VIGour (*N*=120)	0.85	ns	0.08
Relaxed (*N*=132)	2.47	0.014	0.21	[no equiv't]			
[no equiv't]				ANGer (*N*=120)	4.37	<0.001	0.40
Overall mood (*N*=131)	1.63	ns	0.14	TMD	4.51	<0.001	0.41
Median all	2.83		0.25	Median all*	4.38		0.40

* Not including TMD.

The significance level of the Wilcoxon tests for corresponding subscales in NRS-M and $BRUMS_{24}$ are very similar, but effect sizes for all Pre-Post scale differences are small.

[Incidentally, if kurtosis is considered as well as skewness, and the Shapiro-Wilk test is used to check for normality of distribution, only 'Lively' , Pre, Post and Pre-Post score differences are suited to parametric methods of analysis. A paired samples t-test again shows a nonsignificant Pre-Post difference: $t(131) = -0.99$, p=0.32, Cohen's *d* (difference in the means divided by the standard deviation of the difference in means) = 0.09.]

Summary 4 – *Sensitivity of NRS-M to mood and its changes*

Mood

- **Method 1.** Test discrimination $delta_G$ (ability to discriminate between individuals *without* reference to an external criterion) was consistently greater for NRS-M than for $BRUMS_{24}$, both Pre and Post – suggesting that NRS-M has greater sensitivity.
- $Delta_G$ is consistently less for NRS-Post than NRS-Pre.
- **Method 2.** Skewness ÷ SE was consistently positive and almost always numerically greater for $BRUMS_{24}$, indicating that results are probably less dependable than those for NRS-M.
- Skewness ÷ SE was also consistently greater for Post than Pre subscales (except for 'Fatigued').
- **Method 3.** Counts of 'low scores' per individual indicate that NRS-M is more sensitive to all moods, Pre and Post, than $BRUMS_{24}$ (12% vs 32% Pre; 18 vs 43% Post)

Changes in mood

- **Method 1.** $Delta_G$ was consistently greater for NRS-M than for $BRUMS_{24}$ for all subscales.
- **Method 2.** Skewness ÷ SE was good for both scales, if very slightly better for NRS-M.
- **Method 3.** Counts of 'low difference scores' per individual indicate that NRS-M has significantly greater sensitivity to change than $BRUMS_{24}$ for all comparable subscales (73% vs 55%).
- **Method 4.** Numbers of 'no change' Pre-Post differences are significantly more for $BRUMS_{24}$.
- **Method 5.** Comparing Pre and Post subscale median values, $BRUMS_{24}$ appears more sensitive to changes in mood than NRS-M (except for 'Confused'/CONF and 'Lively'/VIG).
- **Method 6.** Significance levels of Wilcoxon tests for Pre-Post differences in corresponding subscales in NRS-M and $BRUMS_{24}$ are very similar, but effect sizes are all small (<0.5).

Objective 5. *To find out whether initial state predicts response when using NRS-M and $BRUMS_{24}$*

Correlations between subscale Pre values and Pre-Post differences were calculated. Disregarding significant values of *rho* <0.3, the highest correlations found were all negative, and between the initial state of a particular scale (e.g. 'Anxious', or 'Relaxed') and changes in the *same* scale. In other words, a high initial value predicted a greater decrease in value, and a low initial value a greater increase in value. This is what would be expected in any case as a 'regression to the mean'. The more

'positive' moods 'Comfortable, 'Relaxed', 'Lively' and VIGour, showed the lowest correlations >0.3, all being <0.6 (**Table 27**).

Table 27. Correlations between initial (Pre) scores and the Pre-Post differences for the same subscale (*rho*>0.3)

NRS subscale	*rho* (Pre-Post)	BRUMS subscale	*rho* (Pre-Post)
Anxious	-0.663	TENSion	-0.646
Comfortable	-0.368	[no equivalent]	
Confused	-0.655	CONFusion	-0.713
Fatigued	-0.631	FATigue	-0.622
Gloomy	-0.680	DEPRession	-0.782
Lively	-0.548	VIGour	-0.468
Relaxed	-0.393	[no equivalent]	
[no equivalent]		ANGer	-0.815
Overall mood	-0.361	[no equivalent]	
TOTAL		TMD	-0.671

It is noteworthy that the 'regression to the mean' which appears here is not supported by the *greater* skewness found consistently for the Post subscales when compared to the Pre subscales (above).

Summary 5 – *Does initial state predict response?*

For both NRS-M and $BRUMS_{24}$ subscales, high initial values predicted greater decreases in value, and low initial values greater increases in value.

This requires further investigation.

Objective 6. *To conduct factor and cluster analyses for NRS-M and* $BRUMS_{24}$

Method 1. Exploratory Factor Analysis (EFA)

An exploratory factor analysis (EFA) was conducted, as described above. Factor loadings after rotation using a significant factor criterion of 0.3 are shown in the **Tables** below.

For NRS-Pre and NRS Pre-Post, two factors were apparent, for NRS-Post, three factors (**Tables 28-30**).

Table 28. Results of exploratory factor analysis for NRS-Pre, showing factor loadings, eigenvalues, percentage of variance, KMO statistics, cumulative percentage variance, percentage of residuals < 0.05, Factor correlations and communalities.

NRS-Pre	Without Overall mood		With Overall mood	
	Factor 1	Factor 2	Factor 1	Factor 2
Comf	0.870		0.813	
Relaxed	0.721		0.762	
Lively	0.624		0.674	
Fatigued	*-0.366*		*-0.389*	
Anx		0.919		0.868
Confus		0.501		0.515
Gloomy	n/a	n/a	n/a	n/a
Overall				
Eigenvalues	2.10	1.60	2.85	1.77
% variance	32.67	11.33	37.52	9.61
KMO	0.669 (0.561-0.701)		0.745 (0.582-0.802)	
Cum % var	44.00		47.13	
% res<0.05	42.9 (9)		46.4 (13)	
Fact correl	-0.446		-0.454	

Note: Communalities <0.3: Confused, Fatigued, Gloomy (both with and without overall mood).

Table 29. Results of exploratory factor analysis for NRS-Post, showing factor loadings, eigenvalues, percentage of variance, KMO statistics, cumulative percentage variance, percentage of residuals < 0.05, Factor correlations and communalities.

NRS-Post	Without Overall mood			With Overall mood		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Comf		0.859		0.919		
Relaxed		0.901		0.874		
Lively			0.503			0.684
Fatigued			*-0.572*		0.304	*-0.405*
Anx	0.798				0.762	
Confus	0.603				0.605	
Gloomy	0.732				0.762	
Overall				0.550		
Eigenvalues	1.94	1.96	1.01	2.57	2.26	1.24
% variance	36.28	15.39	7.10	39.79	13.70	6.76
KMO	0.679			474.17		
Cum % var	58.78			60.24		
% res<0.05	0			0		
Fact correl	-0.245 (F1-F2), -0.271 (F2-F3), -0.281 (F3-F1)			-0.321 (F1-F2), -0.283 (F2-F3), -0.297 (F3-F1)		

Note: Communalities <0.3: Fatigued, Lively (without Overall mood); Fatigued (with Overall mood).

Table 30. Results of exploratory factor analysis for NRS Pre-Post, showing factor loadings, eigenvalues, percentage of variance, KMO statistics, cumulative percentage variance, percentage of residuals < 0.05, Factor correlations and communalities.

NRS Pre-Post	Without Overall mood		With Overall mood	
	Factor 1	Factor 2	Factor 1	Factor 2
Comf	0.983		0.906	
Relaxed	0.860		0.867	
Lively	0.623		0.682	
Fatigued		0.657	0.108	0.656
Anx		0.784	*-0.30*	0.784
Confus		0.513		0.513
Gloomy		0.610		0.610
Overall			0.869	
Eigenvalues	2.25	1.83	2.96	1.87
% variance	35.38	19.68	38.62	18.45
KMO	0.704 (0.634-0.809)		0.762 (0.657-0.815)	
Sphericity	433.522		604.244	
Cum % var	55.06		57.07	
% res<0.05	14.29 (3)		17.86 (5)	
Fact correl	0.229		0.222	

Note: Communalities <0.3: None for 'Extraction' ('Confused' and 'Gloomy' for 'Initial'). Scree plot much more convincing than for either NRS (Pre) or NRS (Post).

In all three analyses, 'Anxious' and 'Confused' (negative valence moods) were consistently associated, as were 'Comfortable' and 'Relaxed' (positive valence moods, which could be considered as a 'Relaxation' factor). 'Gloomy' was associated with the other negative moods in NRS-Post and NRS Pre-Post, and 'Lively' with the positive moods in NRS-Pre and NRS Pre-Post. However, 'Fatigued' was associated with the positive moods (including 'Lively') in NRS-Pre, with the negative moods in NRS Pre-Post, and with 'Lively' in NRS-Post as a third factor (which could be considered as 'Activation').

Method 2. Exploratory Cluster analysis (ECA)

Because sample sizes were not really large enough to conduct a convincing EFA, an exploratory cluster analysis (ECA) was conducted to corroborate it.

Hierarchical cluster analysis was performed by variables on the same data for a range of clusters between 2 and 4, as described above.

The resulting clusters are shown in **Tables 31-33**. A three-cluster solution appeared most appropriate for NRS-Pre, NRS-Post and NRS Pre-Post (although four clusters were also possible for NRS-Pre). Groupings that were stable across solutions with from two to four clusters are also shown.

Table 31. Three-cluster solution for NRS-Pre, with stable clusters across 2-4 clusters.

Without Overall mood		With Overall mood	
Stable clusters	3 Clusters	Stable clusters	3 Clusters
Anxious	Anxious	Anxious	Anxious
Gloomy	Fatigued	Confused	Confused
	Gloomy		Gloomy
Comfortable	Comfortable	Comfortable	Comfortable
Relaxed	Relaxed	Relaxed	Relaxed
Lively	Lively	Lively	Lively
		Overall	*Overall*
	Confused		
			Fatigued

The proximity matrix for NRS-Pre showed nearest neighbours to be 'Comfortable' and 'Relaxed' (4.430 without 'Overall mood', 4.070 with 'Overall mood'), with furthest neighbours 'Comfortable' and 'Gloomy' (43.280 without 'Overall mood', 42.920 with 'Overall mood').

Table 32. Three-cluster solution for NRS-Post, with stable clusters across 2-4 clusters.

Without Overall mood		With Overall mood	
Stable clusters	3 Clusters	Stable clusters	3 Clusters
Anxious	Anxious	Anxious	Anxious
Confused	Confused	Confused	Confused
Gloomy	Gloomy	Gloomy	Gloomy
Comfortable	Comfortable	Comfortable	Comfortable
Relaxed	Relaxed	Relaxed	Relaxed
	Lively	*Overall*	Lively
			Overall
	Fatigued		
			Fatigued

The proximity matrix for NRS-Post showed nearest neighbours to be 'Comfortable' and 'Relaxed' (2.487 with and without 'Overall mood'), with furthest neighbours 'Comfortable' and 'Gloomy' (53.300 with and without 'Overall mood').

Table 33. Three-cluster solution for NRS Pre-Post, with stable clusters across 2-4 clusters.

Without Overall mood		With Overall mood	
Stable clusters	3 Clusters	Stable clusters	3 Clusters
Anxious	Anxious	Anxious	Anxious
Gloomy	Fatigued	Gloomy	Fatigued
	Gloomy		Gloomy
Comfortable		Comfortable	
Relaxed	Comfortable	Relaxed	Comfortable
Lively	Relaxed	*Overall*	Relaxed
	Lively		Lively
			Overall
	Confused		
			Confused

The proximity matrix for NRS Pre-Post showed nearest neighbours to be 'Comfortable' and 'Relaxed' (1.520 with and without 'Overall mood'), with furthest neighbours 'Fatigued' and 'Lively' (11.726 with and without 'Overall mood').

In all three analyses, 'Anxious' and 'Gloomy' (negative valence moods) were consistently associated, as were 'Comfortable', 'Relaxed' and 'Lively' (positive valence moods). 'Confused' and 'Fatigued' were sometimes associated with the negative moods, and sometimes occurred as separate clusters.

The association between 'Comfortable' and 'Relaxed' is thus consistent for both EFA and ECA. For NRS-Pre, 'Comfortable', 'Relaxed' and 'Lively' are associated with both methods of analysis. For NRS-Post, 'Comfortable' and 'Relaxed' are associated, as are 'Anxious', 'Confused' and 'Gloomy', using both EFA and ECA. Best agreement between the two methods is for NRS Pre-Post, with 'Comfortable', 'Relaxed' and 'Lively', and 'Anxious', 'Fatigued' and 'Gloomy' as the main factors/clusters, but lack of agreement about 'Confused'.

Subgroup analysis

For the two subgroups 'CPD' and 'Student', the clusters differed. For example, for NRS Pre-Post, for CPD respondents (*N*=26 because of missing data), the dendrogram indicated a 4-cluster solution as optimal: 'Comfortable'/'Relaxed'/'Lively' – 'Confused' – 'Anxious'/'Gloomy' – 'Fatigued', whereas for the Student respondents (*N*=132), the dendrogram indicated a 3-cluster solution as optimal: 'Comfortable'/'Relaxed'/'Lively' – 'Anxious'/'Confused'/'Fatigued' – 'Gloomy'.

Summary 6 – *Factor and cluster analyses*

Method 1. *Exploratory factor analysis (EFA)*

- For NRS-Pre and NRS Pre-Post, two factors were apparent, for NRS-Post, three factors.
- In all three analyses, 'Anxious' and 'Confused' (negative valence moods) were consistently associated, as were 'Comfortable' and 'Relaxed' (positive valence moods, which could be considered as a 'Relaxation' factor).

Method 2. *Exploratory cluster analysis (ECA)*

- A three-cluster solution appeared most appropriate for NRS-Pre, NRS-Post and NRS Pre-Post (although four clusters were also possible for NRS-Pre).
- In all three analyses, 'Anxious' and 'Gloomy' (negative moods) were consistently associated, as were 'Comfortable', 'Relaxed' and 'Lively' (positive moods). 'Confused' and 'Fatigued' were sometimes associated with the negative moods, and sometimes occurred as separate clusters.
- *Subgroup analysis*. For the two subgroups 'CPD' and 'Student', the clusters differed.

Agreement between the two methods

- The association between 'Comfortable' and 'Relaxed' is consistent for both EFA and ECA.
- For NRS-Pre, 'Comfortable', 'Relaxed' and 'Lively' are associated with both methods of analysis.
- For NRS-Post, 'Comfortable' and 'Relaxed' are associated, as are 'Anxious', 'Confused' and 'Gloomy', using both EFA and ECA.
- Best agreement between the two methods is for NRS Pre-Post, with 'Comfortable', 'Relaxed' and 'Lively', and 'Anxious', 'Fatigued' and 'Gloomy' as the main factors/clusters, but lack of agreement about 'Confused'.

Objective 7. *To investigate common response patterns and styles for both scales, NRS-M and BRUMS$_{24}$*

When a scale such as NRS-M or BRUMS$_{24}$ is used, it is generally assumed that it measures what it purports to measure. But what if the scale itself is inadvertently designed in such a way as to invite particular scores, or styles of scoring? In which case, results may reflect characteristics of the scale itself, or of the respondents who use it, rather than the construct supposedly being measured. To investigate this, two methods were used – calculation of modes (the most commonly occurring score values) and assessment of response style (RS).

Method 1. Modes

Modes were clustered around the numbers on NRS. Those within 1 unit of each decade were counted as that decade (e.g., 79 and 81 were both counted as 80). Some respondents scored in such a way (i.e. with a healthy scattering rather than clustering of scores) that a mode was not found. In some instances where no mode was located automatically by Excel, a near-mode was observed in the data and used instead.

Modes for the two scales, calculated for each *respondent* and then summed, are shown in **Table 34**. Modes for the *subscales* of the two scales are shown in **Table 35**.

Table 34. Modes for NRS-M and BRUMS$_{24}$, Pre and Post – by respondent.

NRS	NRS-Pre	NRS-Post	both	**BRUMS$_{24}$**	BRUMS$_{24}$-Pre	BRUMS$_{24}$-Post	both
no mode	7[a]	4	1	no mode	22	12	2
0	28	39	15	0	64	90	53
[2-8]	1	2	0	1	10	7	0
10	9	15	0	2	3	7	0
[12-18]	0	1	0	3	9	3	1
20	7	2	0	4	11	3	2
30	12	8	0	5	4	3	0

[32-38]	1	0	0	6	3	1	0
40	6	9	0	7	3	0	0
50	19	14	1	8	1	2	0
[52-58]	1	1	0	9	1	1	0
60	17	12	2	10	2	0	0
70	23	15	4	11	1	0	0
75	2	2	0	12	1	0	0
80	9	10	1				
[82-88]	0	1	0				
90	5	6	1				
[92-98]	0	1	0				
100	1	0	0				
SUM	148	142	25	SUM	135	129	58

a. An additional 'no mode' case was observed to use values around 90 most commonly.

Mode was most often 0, for both NRS-M and $BRUMS_{24}$. Next most common mode was 70 for NRS-Pre, and 70 or 10 for NRS-Post, with 50 in third place for both NRS-Pre and NRS-Post. Except perhaps for $BRUMS_{24}$ mode 4 (11 Pre, 3 Post), the modes for the Pre and Post scales are quite similar, as assessed by their correlation (for NRS-M, *rho*=0.899, p<0.001; for $BRUMS_{24}$, *rho*=0.799, p=0.001).

For each scale, respondents appear likely to score the Pre and Post versions in quite similar ways, but not so similarly as to mask changes that are perceived.

There is an overlap of 25 between NRS-Pre (16.9%) and NRS-Post (17.6%) modes, and of 58 between $BRUMS_{24}$-Pre (43.0%) and $BRUMS_{24}$-Post (45.0%) subscale modes. Thus the degree of overlap is similar for Pre and Post versions of the same measure, but more for $BRUMS_{24}$ than NRS-M. This suggests that there is less fixity of mode for NRS-M than for $BRUMS_{24}$, or in other words that NRS-M may reflect changes in mood in a less rigid, more sensitive way than $BRUMS_{24}$.

Table 35. Modes for NRS-M and $BRUMS_{24}$, Pre and Post – by subscale.

NRS subscale	NRS-Pre	*N*	NRS-Post	*N*	both	BRUMS subscale	BRUMS-Pre	*N*	BRUMS-Post	*N*	both
Anxious	0	38	0	62	24	TENSion	0	42	0	59	29
Comfortable	70	25	80	33	n/a	[no equiv't]					
Confused	0	37	0	46	23	CONFusion	1	24	0	54	(15)[a]
Fatigued	70	24	50	29	n/a	FATigue	12	17	4	19	n/a
Gloomy	0	45	0	74	31	DEPRession	0	66	0	90	56
Lively	60	27	50	26	n/a	VIGour	8	18	8	17	3
Relaxed	70	28	70	30	9	[no equiv't]					
[no equiv't]						ANGer	0	75	0	92	57
Overall mood	70	31	70	35	11	[no equiv't]					
Mode all	70		0			TMD	10	6	5	8	n/a

a. Taking the mode of BRUMS-Pre CONF as 0.

When considering modes by *subscale* rather than by *respondent*, it appears, for example, that more respondents indicated zero levels of 'Anxious'/TENS and 'Gloomy'/DEPR than any other amount, and more did so towards the end of the teaching session than early on during it. Findings were similar for

'Confused'. In contrast, the most commonly reported level of 'Fatigued'/FAT dropped, but numbers reporting that level did not change greatly. For 'Comfortable', both levels and numbers of respondents reporting those levels increased somewhat, whereas for 'Relaxed' and 'Overall mood', there was little change.

Those subscales with least change in either level or count of the mode would be those most likely to be insensitive to change, possibly because respondents feel forced in some way to complete them in particular ways just because of how the subscales are presented.

Table 36 shows the overlaps between subscales with identical modes in Pre and Post versions.

Table 36. Overlaps between subscales with identical modes in Pre and Post versions of NRS-M or $BRUMS_{24}$.

NRS subscale	Overlap (*N*)	NRS-Pre	NRS-Post	BRUMS subscale	Overlap (*N*)	BRUMS-Pre	BRUMS-Post
Anxious	24	63.2%	38.7%	TENSion	29	69.0%	49.2%
Confused	23	62.2%	50.0%	CONFusion	15	62.5%	27.8%
Gloomy	31	68.9%	41.9%	DEPRession	56	83.3%	61.1%
Relaxed	9	32.1%	30.0%	*VIGour*	3	16.7%	17.6%
Overall m	11	35.5%	31.4%	*ANGer*	57	76.0%	62.0%
Medians	23	62.2%	38.7%		29	69.0%	49.2%

The degree of overlap is similar for Pre and Post versions of 'Relaxed' and 'Overall mood', but quite different for the other NRS-M subscales. For $BRUMS_{24}$ there is only similar overlap in Pre and Post versions for VIGour. Median *N* and overlap percentages suggest again that there is slightly less fixity of mode for NRS than for $BRUMS_{24}$, or in other words that NRS-M may reflect changes in mood in a less rigid, more sensitive way than $BRUMS_{24}$.

However, results are less clear-cut when individual subscales are considered and compared for the two measures. Mode overlaps for $BRUMS_{24}$ subscales DEPR and ANG suggest that these two $BRUMS_{24}$ subscales are those most likely to provide results for mood which are influenced by the *way* the subscale is scored (possibly regardless of who is doing the scoring), rather than being accurately reflected in the scores themselves.

Use of the most frequently occurring mode (0) across the two measures

In addition to exploring patterns of response that may be similar in Pre and Post versions of NRS-M and $BRUMS_{24}$, similarities between how respondents completed NRS-M and $BRUMS_{24}$ subscales were investigated.

23 of those with '0' mode for NRS-Pre (82.1%) also scored $BRUMS_{24}$-Pre (35.9%) with '0' mode, but only 1 of those using 'no mode' for NRS-Pre (14.3%) did so for $BRUMS_{24}$-Pre (4.5%).

31 of those with '0' mode for NRS-Post (79.5%) also scored $BRUMS_{24}$-Post (34.4%) with '0' mode, but only 1 of those using 'no mode' for NRS-Post (25.0%) did so for $BRUMS_{24}$-Post (8.3%).

Thus '0' mode scores for NRS-M (particularly for NRS-Pre) are good predictors of '0' mode scoring for $BRUMS_{24}$, but not vice versa.

Method 2. Response style (RS)

Response style (RS), or the tendency to respond to questionnaire items independently of item content, is a source of contamination in questionnaire responses, and therefore threatens the validity of conclusions (Baumgartner & Steenkamp 2001; Weijters et al. 2008). The most commonly reported RSs are: (1) acquiescence RS (ARS), or 'yea-saying'; (2) disacquiescence RS (DARS), or 'nay-saying'; (3) net acquiescence RS (NARS), or directional bias (more acquiescence than disacquiescence); (4) extreme RS (ERS), or the tendency to endorse the most extreme response categories; (5) response range (RR), or the tendency to use a narrow OR wide range of response categories around the mean response; (6) midpoint responding (MRS), or the tendency to use the middle scale category regardless of content; (7) and noncontingent responding (NCR), or the tendency to respond to items carelessly, randomly or nonpurposefully. Additional RSs include, for example, non-extreme RS (NERS) (Wetzel et al. 2013) sometimes known as a 'mild RS' (Van Vaerenbergh & Thomas 2013), and non-midpoint responding (NMRS).

Response styles tend to be stable over time, but may differ with age, gender and education level, as well as mood, fatigue and cognitive load (Weijters et al. 2010a; Van Vaerenbergh & Thomas 2013). ERS has been found to be positively related to personality measures such as intolerance of ambiguity, preference for simple thinking and decisiveness, and personality traits such as extraversion and conscientiousness. Midpoint scoring (MRS) may attract truly neutral/indifferent respondents (being neither positive nor negative), but also ambivalent respondents. However, most RS remains unexplained, and in particular how RS relate to central personality traits. One person who has explored this is Elke Cabooter, who has examined the difference between 'promotion-focused' people (more open to risk) and 'prevention-focused' people (more vigilant, and concerned to avoid danger or loss), and shown that the former show higher levels of ERS, with the reverse being true for MRS (Cabooter 2010).

However, it is important to note that personality measures might themselves be contaminated with RS (Van Vaerenbergh & Thomas 2013). Clearly a major methodological difficulty in the present context is that mood (ratings) may both affect RS and be impacted by RS. Cabooter has proposed studies to investigate the effect of mood on RS and to compare results using VAS vs Likert scales (Cabooter 2010).

Furthermore, data obtained with different formats are not necessarily compatible. In particular, labelling all response categories, or adding a midpoint, leads to lower levels of ERS (Weijters et al. 2010b). The implication in the present context is that ERS might be greater for NRS-M than for $BRUMS_{24}$ (but see below, p 40).

<u>Response style – analysed by respondents</u>

The definitions for RS here are shown in **Tables 37** (for NRS-M) and **38** (for $BRUMS_{24}$).

Table 37. Definitions of NRS-M response styles, with thresholds for 'high' (and 'low') scores.

NRS-M	by subscale	Threshold (sum by respondent)
ERS	≤10 or ≥90	≥6[a]
NERS	>10 and <90	≥6.5[a] (≈ ERS ≤3)
MRS	≥40 and ≤60	≥7[a]
NMRS	<40 or >60	≥8[a]
RR	standard deviation of responses (SD)	>40 (Pre); >46 (Post)[b]
NCR	average difference between pairs of ratings with *rho* >40% in complete sample (van Rosmalen et al. 2010)	high: >40 (Pre); >36 (Post)[b] low: <10 (Pre); <12 (Post)[b]

a. Summed by respondent; b. SDs for each respondent were ranked and graphed, and the resulting S-shaped curve compared with its linear trendline. A visual estimate was made of where either end of the curve diverged from the line.

Table 38. Definitions of $BRUMS_{24}$ response styles, with thresholds for 'high' (and 'low') scores.

$BRUMS_{24}$	by subscale items	Threshold (sum by respondent)
ERS	0 or 4	≥17[a]
NERS	1, 2 or 3	≥6.5[a] (≈ ERS ≤3)
MRS	Not found in this sample	n/a
NMRS	Not appropriate for this sample	n/a
RR	standard deviation of responses (SD)	high: >5.6 (Pre); >5.4 (Post)[b] low: <2.2 (Pre); <2 (Post)[b]
NCR	Not calculated	n/a

a. Summed by respondent; b. SDs for each respondent were ranked and graphed, and the resulting S-shaped curve compared with its linear trendline. A visual estimate was made of where either end of the curve diverged from the line.

Our data was not well suited to assessment of ARS, DARS or NARS, so these were not considered for analysis.

Numbers (percentages) of respondents exhibiting particular RSs based on these criteria are shown in **Tables 39 (for NRS-**M) and **40** (for $BRUMS_{24}$).

Table 39. Numbers (percentages) of respondents exhibiting particular response styles for NRS-M.

Response style	NRS_Pre (*N*=148)	%	NRS_Post (*N*=142)	%	NRS (Both) (*N*=132)	%
ERS	15	10.1	17	12.0	6	4.5
NERS	114	77.0	97	68.3	97	73.5
MRS	10	6.8	7	4.9	1	0.8
NMRS	22	14.9	29	20.4	8	6.1
RR	21	14.2	25	17.6	11	8.3
NCR	8	5.4	7	4.9	0	0.0

Clearly, for NRS-M most respondents exhibited a 'mild response style' (or NERS), and otherwise numbers of those who exhibit a particular RS *consistently* in NRS_pre and NRS_post are small.

Table 40. Numbers (percentages) of respondents exhibiting particular response styles for $BRUMS_{24}$.

Response style	BRUMS_Pre (*N*=135)	%	BRUMS _Post (*N*=129)	%	BRUMS (Both) (*N*=121)	%
ERS	19	14.1	23	17.8	8	6.6
NERS	20	14.8	20	15.5	9	7.4
RR	13	9.6	13	10.1	3	2.5

The pattern of response is very different for $BRUMS_{24}$ than for NRS-M, with approximately equal numbers of consistent ERS and NERS respondents, and relatively fewer showing high RR. The numbers and percentages of respondents exhibiting ERS are marginally higher for $BRUMS_{24}$ than for NRS-M, a result opposite to that anticipated (above, p 38).

Numbers (percentages) of those who exhibited particular response styles for both NRS-M and $BRUMS_{24}$ are shown in **Table 41**.

Table 41. Numbers (percentages) of those exhibiting particular response styles for both NRS-M and $BRUMS_{24}$.

Response style	Pre (Both) (*N*=134)	%	Post (Both) (*N*=130)	%	Pre & Post (Both) (*N*=121)	%
ERS	7	5.2	6	4.6	1	0.8
NERS	18	13.4	18	13.8	8	6.6
RR	2	1.5	5	3.8	1	0.8

These results suggest that, at least in this sample, RSs that are consistent across both Pre and Post versions of NRS-M and $BRUMS_{24}$ do not threaten the validity of conclusions drawn from responses to the two scales, but that RS may impact conclusions drawn from one or the other – and more so for $BRUMS_{24}$.

Overlapping response styles (by respondent)

The categories of RS used here are not necessarily mutually exclusive. **Table 42** shows numbers of NRS-M overlaps between pairs of different response styles for this sample, with the percentages of the total numbers of each response style in the pair and the correlations between them.

Table 42. Numbers of NRS-M overlaps between pairs of different response styles for this sample, with the percentages of the total numbers of each response style in the pair (in that order) and the correlations between them.

Response style	NRS_Pre	R^2	NRS_Post	R^2	NRS (Both)
ERS & RR	13 (87%; 62%)	0.752	15 (88%; 60%)	0.598	4 (67%; 36%)
ERS & NMRS	9 (60%; 43%)	0.364	12 (71%; 41%)	0.390	1 (17%; 13%)
NMRS & RR	9 (41%; 43%)	0.587	18 (62%; 72%)	0.560	4 (50%; 36%)
NCR & ERS	3 (38%; 20%)		2 (29%; 12%)		0
NCR & RR	2 (25%; 10%)		2 (29%; 8%)		0
NCR & NMRS	1 (13%; 5%)		0		0

The greatest overlaps are between ERS and RR, ERS and NMRS and NMRS and RR. Given the prevalence of ERS in the literature, the partial redundancy of this with the other two measures of RS (NMRS and RR), and the relative complexity of calculating NCR, ERS and MRS will be considered primarily in further analysis.

Response style and mode (by respondent)

As would be expected, ERS for NRS-M and $BRUMS_{24}$ occurred most frequently for mode 0. NERS occurred most frequently for NRS-M for non-extreme modes 50, 60 and 70 (but also – surprisingly – for mode 0 in NRS-Post). For $BRUMS_{24}$, NERS occurred most frequently for non-modal scoring. Results for MRS and NMRS were as expected, the former most frequent for mode 50, the latter for mode 0. RR greater than the threshold set occurred most frequently for mode 0, but not frequently for high number modes. NCR (random scoring) occurred most frequently for mode 0 – probably an artefact of mode 0 itself occurring most often.

Subscale scoring by high ERS respondents

What subscales (or, for $BRUMS_{24}$, subscale items) do respondents with an ERS score high or low?

For both scales, 'negative' valence moods are revealed far less, and 'positive' valence moods more, by ERS respondents, as shown in **Figures 3** and **4**.

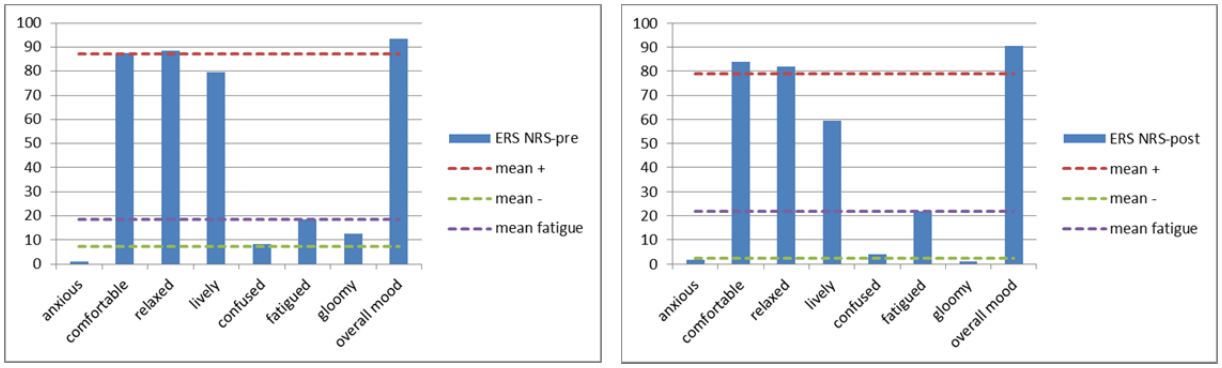


Fig 3. Scoring of subscales by ERS respondents in NRS-M.

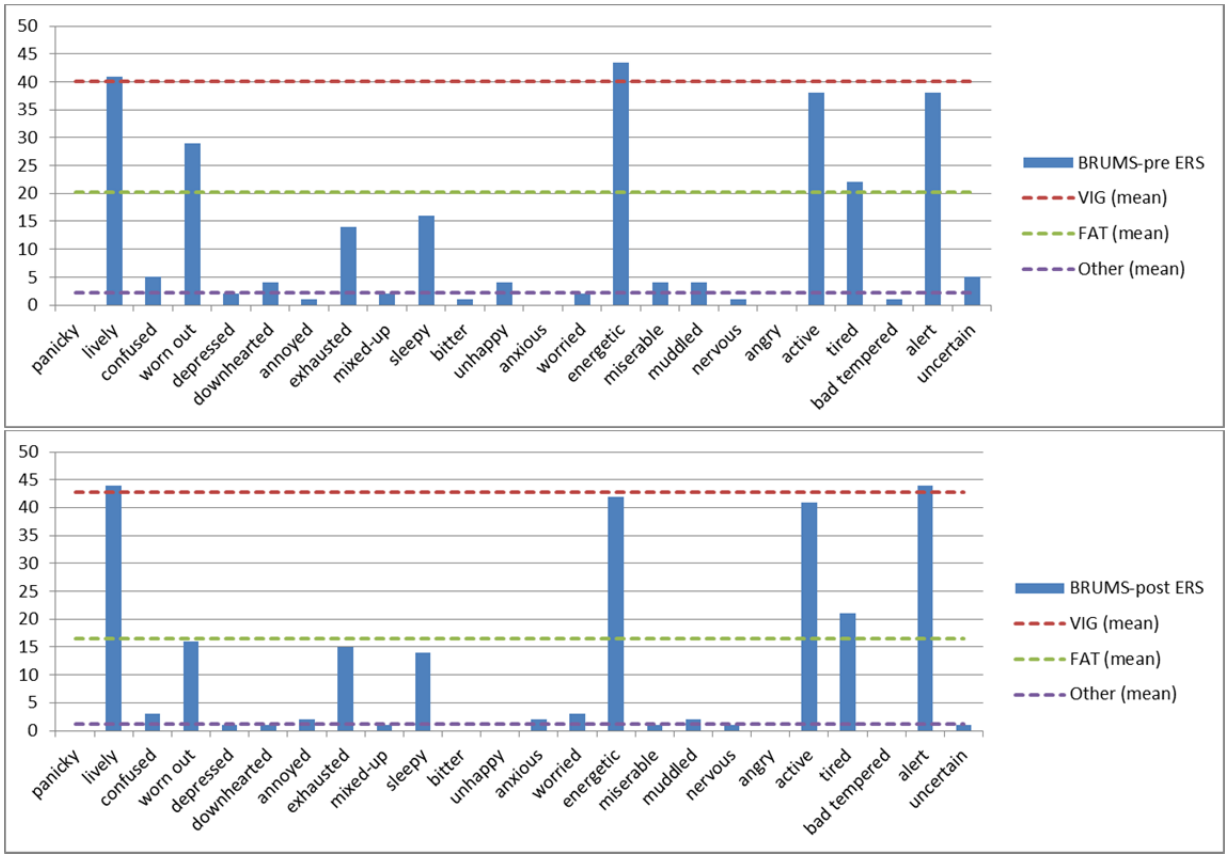


Fig 4. Scoring of subscales by ERS respondents in $BRUMS_{24}$.

The important finding here is that ERS respondents tend not to reveal inner mood so much as their more (socially acceptable?) 'Lively'/VIG and 'Fatigued'/FAT status. In other words, the 'positive' ('Comfortable', 'Relaxed', 'Lively', 'Overall mood') are emphasised by ERS respondents, the 'negative' ('Anxious', 'Confused', 'Gloomy') de-emphasised (with 'Fatigued' somewhere between positive and negative). For NERS respondents, there is less difference between positive, negative and fatigued mood states; 'negative' moods are revealed far less by ERS than by NERS respondents (results not shown).

The Kruskal-Wallis test shows these visually obvious differences between the 'Lively'/VIG, 'Fatigued'/FAT and Other scores are significant (**Table 43**).

Table 43. Significance of differences between the 'Lively'/VIG, 'Fatigued'/FAT and Other scores for ERS respondents.

Scale	H(2)	Significance
NRS-Pre	78.467	p<0.001
NRS-Post	87.337	p<0.001
BRUMS-Pre	16.211	p<0.001
BRUMS-Post	16.436	p<0.001

In addition, **Table 44** shows significant correlations between individual subscale scores and the degree of ERS or MRS for which Spearman's *rho*>0.6 (ERS being considered as an ordinal variable). No correlations with *rho*>0.6 were found for $BRUMS_{24}$.

Table 44. Correlations with *rho*>0.6 between individual subscale scores and the degree of ERS or MRS.

Scale	RS	Subscale	*rho*
NRS-Pre	ERS	Anxious	-0.613
NRS-Post	ERS	Anxious	-0.685
		Confused	-0.637
		Gloomy	-0.617
	MRS	Overall mood	-0.631

Again, ERS respondents would be less likely to score these 'negative' moods highly. In addition, MRS respondents are less likely to score ('positive') 'Overall mood' highly (perhaps a rather obvious finding!).

Subsidiary response styles (SRS)

In addition to the usually recognised response styles, there may be some that are particular to NRS format, not shared with the Likert-style format of $BRUMS_{24}$:

1. Low SRS — Tending to score <50 irrespective of subscale scored
2. High SRS — Tending to score >50 irrespective of subscale scored
3. Low decade SRS — Tending to score *1, *2 or *3 irrespective of decade and subscale scored, rather than *0
4. Mid decade SRS — Tending to score *4, *5 or *6 irrespective of decade and subscale scored, rather than *0
5. High decade SRS — Tending to score *7, *8 or *9 irrespective of decade and subscale scored, rather than *0.

Numbers (%) in this sample exhibiting a predominant SRS across *both* NRS-Pre and NRS-Post (*N*=132):

1. Low SRS — 2 (1.5%)
2. High SRS — 0
3. Low decade SRS — 1 (0.8%)
4. Mid decade SRS — 7 (5.3%)
5. High decade SRS — 0

Thus more respondents score using a mid decade SRS than with other SRSs. A simple explanation for this is that these respondents may have been under the impression that they had to mark the 0-100 line *between* rather than *at* the intervening numerical markers. It is not known if those who score using the other SRSs have particular associated characteristics (see Discussion).

Problems of reverse scoring and persona

Reverse scoring

Given the reasonable correlations between comparable subscales in the two measures, scores should generally be in the same direction – e.g. the same participant should score both low 'Fatigued' and low FATigue, or high 'Lively' and high VIGour. Counts were made of 'low' and 'high' scores, taken as ≤20 or ≥80 for NRS-M subscales, or ≤3.2 or ≥12.8 for $BRUMS_{32}$ subscales. Concurrent high NRS-M AND low $BRUMS_{24}$ scores, or low NRS-M AND high $BRUMS_{24}$ scores, are shown in **Table 45**.

Table 45. Numbers of reverse scores for comparable subscales in NRS-M and $BRUMS_{24}$, with 'low' defined as ≤20 or ≤3.2, and 'high' as ≥80 or ≥12.8, respectively.

NRS-M/ $BRUMS_{24}$	Anxious/TENS		Confused/CONF		Fatigued/FAT		Gloomy/DEPR		Lively/VIG	
	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi
Pre	0	0	0	1	0	0	0	3	1	0
Post	0	1	0	0	1	0	0	0	0	1

Very little extreme reverse scoring is evident – apart perhaps from 'Gloomy' in relation to DEPR.

If the criteria for reverse scoring are relaxed, counts being made of scores ≤30 or ≥70 for NRS-M subscales, or ≤4.8 or ≥11.2 for $BRUMS_{32}$ subscales, more reverse scores appear (**Table 46**), but even so, *consistent* reverse scoring (Pre and Post) is infrequent.

Table 46. Numbers of reverse scores for comparable subscales in NRS-M and $BRUMS_{24}$, with 'low' defined as ≤30 or ≤4.8, and 'high' as ≥70 or ≥11.2, respectively.

NRS-M/ $BRUMS_{24}$	Anxious/TENS		Confused/CONF		Fatigued/FAT		Gloomy/DEPR		Lively/VIG	
	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi	hi/lo	lo/hi
Pre	0	3	0	4	1	1	0	10	1	0
Post	0	3	0	2	1	1	0	2	1	0

Correlation between 'Gloomy' and DEPR Pre subscales is comparatively low (**Table 5**), suggesting that, in addition to reverse scoring, some participants may interpret the two differently. There may be a low incidence of reverse scoring for the other subscales, with NRS-M scored *low* instead of high for 'Anxious', 'Confused' and 'Depressed', and perhaps *high* instead of low for 'Lively' (with reverse scores in either direction possible for 'Fatigued').

An alternative interpretation might be that scoring to the left along a relatively undifferentiated line as in NRS-M might in itself reflect an underlying negative emotional state and relative right cerebral hemisphere activation (He et al. 2010). 'Pseudoneglect' (bisecting a horizontal line to the left of its true centre) is in any case not uncommon (Hausmann et al. 2002), and as He et al. have shown, may indeed be associated with higher levels of anxiety or depression (certainly in those with generalised anxiety disorder or treatment-resistant depression). However, no data was collected on the psychological health of respondents in the current study, nor on their handedness, tendency to

right-left confusion or (for women) their time of the month in the menstrual cycle, any of which may also influence spatial awareness (He et al. 2010). Furthermore, the pseudoneglect interpretation does not explain why results are not consistently biased across all NRS-M subscales, although of course it could be masked for the negatively skewed subscales (see above). An intriguing possibility is that a consistent 'Low SRS' or 'Low decade SRS' may also be associated with relative right cerebral hemisphere activation, but, to our knowledge, this has never been investigated.

Persona scoring

In certain situations, such as questionnaires for job applicants, a degree of 'impression management' would be expected, with respondents emphasising their socially desirable characteristics, even if not consciously (Lorr 1989; Zelenbrz 2005). Even in the present context, where no judgment was being passed and responses were anonymised, fewest respondents were willing to admit (or maybe were aware of) socially less acceptable emotions such as 'Anxious'/TENS, ANG or 'Gloomy'/DEPR. It was easier for most respondents to report levels of 'Fatigued'/FAT or 'Lively'/VIG. 'Comfortable', 'Relaxed' and 'Overall mood' (all with a 'positive' valence) were likewise not problematic. From a few known cases observed while completing NRS-M, it is tempting to speculate that respondents who have attended a psychotherapy or other self-awareness training course (and/or who were themselves in psychotherapy) would complete these scales quite differently. Those willing to state they were 'Confused'/CONF were fewer than those reporting 'Fatigued'/FAT or 'Lively'/VIG, but still somewhat more than those reporting 'Anxious'/TENS or 'Gloomy'/DEPR. Persona scoring as an issue may require further study.

Summary 7 – *Common patterns and response styles*

- Do results reflect characteristics of the scale itself, of respondents using it, or the construct supposedly being measured?

Method 1. *Modes*

- Respondent mode was most often 0, for both NRS-M and BRUMS24. Next most common mode was 70 for NRS-Pre, and 70 or 10 for NRS-Post, with 50 in third place for both.
- Less respondent overlap between Pre and Post modes for NRS-M than $BRUMS_{24}$ suggests that NRS-M may be more sensitive to mood changes.
- Subscale mode was 0 for 'Anxious'/TENS and 'Gloomy'/DEPR, 70 or 80 for 'Comfortable' and 'Relaxed'.
- Mode '0' scores for NRS-M (particularly for NRS-Pre) are good predictors of mode '0' scoring for $BRUMS_{24}$, but not vice versa.

Method 2. *Response styles (RS)*

- Most NRS-M respondents exhibited a 'mild response style' (or non-extreme RS, NERS).
- Numbers of those who exhibited a particular RS consistently in NRS_pre and NRS_post are small (<9%, or <6% for subsidiary response styles, SRS).
- *Consistent* RSs across *both* NRS-M and $BRUMS_{24}$ (0.8%) do not threaten the validity of conclusions drawn from responses to the two scales.
- ERS respondents reveal (score) positive valence moods more, and negative valence moods far less (a significant difference). Thus they emphasise positive moods and de-emphasise negative moods.
- Consistent reverse scoring (Pre and Post) is infrequent. Persona scoring as an issue requires further study.

Objective 8. *On a small sub-sample (N=13), to explore whether there are any significant relationships between response patterns and emotional intelligence characteristics at baseline*

Associations between ERS and emotional intelligence scores from the TEIQue scale (Petrides 2009) were assessed, with both variables treated as nominal. Resulting values of *phi* and Goodman and Kruskal's *tau* suggest that NRS-Pre ERS is positively associated with a high Global TEIQue score (*phi*=0.843, p<0.002), and NRS-Post ERS with a high TEIQue Wellbeing score (*phi*=0.837, p=0.004). BRUMS-Post ERS, however, appears positively associated with a high TEIQue Self-motivation score score (*phi*=0.843, p<0.002).

These tentative findings suggest that it may be worthwhile investigating whether emotional intelligence affects RS when scoring mood as well as the mood scores themselves.

Summary 8 – *Extreme response pattern (ERS) and trait emotional intelligence using the TEIQue-SF*

- NRS-Pre ERS is positively associated with a high 'Global' emotional intelligence score.
- NRS-Post ERS is associated with a high 'Wellbeing' emotional intelligence score.
- These tentative findings require further investigation.

Objective 9. *On a second small sample (N=16), to check whether users find it quicker and easier to use NRS-M or BRUMS$_{24}$ and to ascertain their preferences for one or the other*

A convenience sample of 16 was recruited from friends, relatives and those passing the laboratory where we were undertaking research. They were asked to complete one and then the other of NRS-M and BRUMS$_{24}$ (order balanced across the group). How long it took to complete each scale was recorded, and then they were asked for their comments in a brief nonstructured interview. The whole procedure usually took less than eight minutes. Ages were recorded, if known, or estimated, if not asked. Verbal responses were written down, not electronically recorded.

Ten women and 6 men were recruited, aged an average of 44.5 (SD 19.0) and 43.5 (SD 15.2) years, respectively (median age for the group being 49). **Table 47** shows the median times taken to complete the two measures for men, women and all respondents, in seconds.

Table 47. Median times taken to complete the two measures for men, women and all respondents, in seconds, with the ratio between them.

Gender	Age (y) (mean, SD)	NRS-M (median)	Range (s)	BRUMS$_{24}$ (median)	Range (s)	Ratio (N/B) (median)
Women	44.5, 19.0	44	34-82	80	65-216	0.51
Men	43.5, 15.2	41	26-49	74.5	37-103	0.59
All	44.1, 17.1	41	26-82	80	37-216	0.51

Users completed NRS-M in markedly less time than they did BRUMS$_{24}$. Seven NRS-M responses (43.8%) took between 20 and 30 seconds, and eight BRUMS$_{24}$ responses (50.0%) between 60 and 90 seconds. Moreover, in this small sample, men appear to complete the scales marginally more quickly than women.

Table 48 shows the median times taken to complete the two measures for those < or ≥ the group median age (49).

Table 48. Median times (seconds) taken to complete the two measures for those < or ≥ the group median age (49).

Age	NRS-M (median)	Range (s)	BRUMS$_{24}$ (median)	Range (s)	Ratio (N/B) (median)
< median age	39	26-82	70	37-156	0.56
≥ median age	47	31-64	90.5	63-216	0.48
All	41	26-82	80	37-216	0.51

In addition, younger people appear to complete the scales – particularly BRUMS$_{24}$ – quicker than older people, although this difference is not significant (using the Mann-Whitney test for two independent samples).

Summary of results from brief unstructured interviews (as well as in volunteered comments)

16 brief, informal interviews were conducted, with further comments volunteered by four participants in an ongoing EEG/HRV study in which these measures of mood are being used. Preferences and opinions were as follows:

Preferences

NRS-M: Four respondents preferred NRS-M, either because they found it easier, or because they just liked numbers.

BRUMS$_{24}$: Eight preferred BRUMS$_{24}$, finding it easier and (for four respondents) 'more precise'. Two of these felt that they had to think more or longer with NRS-M. One was happy that 'there was more to do' with BRUMS$_{24}$, and another acknowledged that NRS-M was 'simpler'.

Three respondents expressed no preference (e.g. 'What's the difference? Line or box – it's the same thing'), although one of these liked 'having the actual scale' (of NRS-M). Another who expressed no preference considered BRUMS$_{24}$ 'more descriptive and more accurate'.

Sensitivity

One respondent suggested that NRS-M percentages 'may vary more' with time, and so may be 'more appropriate for before/after changes'. Two of those who preferred BRUMS$_{24}$ acknowledged that NRS-M was 'probably more sensitive' or that it gives 'a more accurate perspective'. One who preferred NRS-M considered herself 'a very visual person, so [found it] more subtle, [with] more gradation'. Another observed a 'floor effect' when completing BRUMS$_{24}$, with many items scored 'not at all' just because of the way the questionnaire is structured. However, as noted above, several respondents considered BRUMS$_{24}$ to be more precise or accurate.

Language

The terms in both measures were generally considered 'very' or 'quite' easy, even by those for whom English was not their primary language. One found it 'quite clarifying' to have a variety of terms for a similar construct in BRUMS$_{24}$, but another that 'Some questions are a bit odd'.

The distinction between mood and emotion was highlighted by one respondent, who observed that terms like 'Lively' and 'Alert' 'are more about energy than mood, mood [being] something that takes you over, [whereas] some of these are emotions'. On 'Confused', the same respondent stated:
'I can't answer that one because I don't know what I'm supposed to be confused about; … anxious – but about what?'. Later, she said: 'I don't know what 'mixed up' means'.

Someone familiar with the valence/arousal model of mood found 'Confused' awkward because it was hard to position 'on an arousal axis' without further information (e.g. on its relationship with anger or depression).

One of our EEG/HRV study participants commented scathingly on two BRUMS$_{24}$ items: 'How can I be *a bit* exhausted?' and 'How can I feel energetic if I can't move?'.

Number

Attitudes to number varied greatly. One respondent, for example, was heard to say to herself 'don't be too greedy' when scoring one subscale as 90 rather than 100. Another, who was more used to 0-10 scales, found 1-30-40 … 'a bit strange'. One researcher commented that with the 0-100 scale results might be normally distributed, whereas you 'couldn't argue this for [$BRUMS_{24}$]'. And a less science-oriented respondent that 'I've no idea how to do numbers'.

Dislikes or difficulties

NRS-M: One respondent was aware of the danger of reverse scoring with NRS-M ('I feel I wanted to tick one end, then realised it should be the other'). Another commented that he had to check 'the negative and positive [anchor phrases] … one vs the other as they're quite far [apart, and] … there were no words all the way along'. Another (with a history of mild right-left confusion) said 'I don't know … I've no idea how to do numbers – I've just gone round the middle'.

$BRUMS_{24}$: One found $BRUMS_{24}$ 'more clunky, … almost a bit confusing', another found it 'too heady, and it's repetitive – not many nice feelings, so it doesn't put you in a good frame of mind'. Another commented that it was 'a little harder to understand visually – although obvious when you read the instructions. [It] looks like you're meant to write in the boxes as the numbers are not centred. [It] looks more complex than it is'. This same respondent 'thought [some] words were similar, and couldn't help checking back to see what I'd put for similar words as didn't want to contradict myself … Tried to be consistent – almost automatically – started doing it before I realised'.

One criticism of $BRUMS_{24}$ was that there are 'a lot of words to read through', and indeed some respondents spent almost as much time reading through the initial instructions as completing NRS-M (median 33 vs 38 seconds, although time taken to read the instructions was only recorded for N=4).

A telling criticism of such questionnaires in general was the following: 'I hate doing these sorts of things – it's very stupid; ... I'm reacting to doing the questionnaire; what level do you want me to go to?'.

Discussion

This small qualitative nested study highlights that how long it takes to complete a questionnaire may have little to do with how respondents feel about it. Differences in time taken showed a similar pattern for the whole group, but attitudes to the two measures were very individual. Attitudes may depend, for example, on whether respondents are more comfortable with numbers or with words, whether they have a need to check that they are doing the 'right' thing in some way (or the opposite, if they dislike Procrustean conformity), whether they are more scientific /rational/'left-brained' or more artistic/intuitive/'right-brained' in their approach, and so on. Researchers may prefer one method, respondents another, and this in itself could affect how they score their mood on these measures.

Summary 9 – *Do users find it quicker and easier to use NRS-M or BRUMS$_{24}$, and what are their preferences?*

- Users (and younger men in particular) completed NRS-M in markedly less time than they did BRUMS$_{24}$.
- More users preferred BRUMS$_{24}$, but opinions were divided on which scale was more accurate.
- In general the terms used were readily understood, even by those for whom English was not their primary language.
- Personal characteristics and attitudes to measures of mood may affect how moods are scored.

Discussion – some advantages and disadvantages of the two scales

In this study, NRS-M and BRUMS$_{24}$ were used to assess mood and its changes during a practical teaching seminar, with results that could potentially be used to further improve teaching methods and which also suggest ways of measuring outcomes in response to acupuncture treatment.

Measuring subjective states such as mood or emotion is not straightforward, and was considered reductionist and superficial by at least one respondent in the present study. There is also a fundamental lack of agreement on which are the most important moods (or emotions) to measure. A mood's relevance will vary in different studies depending on context, treatment, population and objectives. For example, it is probably naïve to expect respondents to reveal much about their personal state in a research context, whereas they might be more willing to do so if given time to establish trust in their practitioner in a therapeutic setting. Furthermore, there is not always consensus on which terms can be construed as facets of the same construct – 'Exhausted' and 'Sleepy' are both taken to be aspects of FATigue in BRUMS$_{24}$, for example, but elsewhere 'vital exhaustion' has been shown to differ from fatigue and sleepiness and – not surprisingly – to be associated with a whole slew of other negative state affects, the degree of association being modified by temperament (Heponiemi et al. 2005). Mood-related states such as boredom, comfort and relaxation may also need to be considered – they may be appropriate in some settings but not others, for example.

Confounding factors in mood measurement can include response style, reverse scoring, what we have termed 'persona scoring', and inherent skewness when subscales are asymmetric. Indeed, mood itself may affect the way a person scores a horizontal NRS, irrespective of what the NRS purports to measure (He et al. 2010). Vertical scales may be preferable in those with hemispatial neglect such as post-stroke patients (Arruda et al. 1999), but in a relatively healthy population – as in the present study – use of a horizontal scale should not be overly problematic. A particular confounding factor in comparing test-retest reliability for NRS-M and BRUMS$_{24}$ in the present study was that Pre and Post versions of the former were printed in different orders, whereas for BRUMS$_{24}$

they were in the same order. This may have affected our findings, although we hope not in any major way.

Although neither scales is difficult to use, one advantage of NRS-M over a multi-item questionnaire is that it uses fewer words, and if these are familiar to respondents, it could be easier to complete than the POMS or $BRUMS_{24}$.

There will always be people who prefer numbers and others who prefer words, but the incidence in the general population of those who have real difficulties with number (as in dyscalculia) is usually considered to be somewhat less (3-6%) (Shalev et al. 2001) than that of those who have severe problems with words (as in dyslexia) (5-17.5%) (Siegel 2006; Lagae 2008) – although the incidence of 'arithmetic dysfluency' may be comparable, at least in children (Reigosa-Crespo et al. 2012). Although those with mathematical learning difficulties may struggle to some extent with number line estimation tasks (Van't Noordende et al. 2015) akin to those involved in NRS-M, overall there may be a slight advantage in using NRS-M rather than $BRUMS_{24}$ for those with non-severe learning difficulties.

One issue with $BRUMS_{24}$ is its focus on pathology. In acupuncture teaching and practice, the emphasis is often more on health as a positive state of wellbeing rather than simply the absence of disease or disorder. An advantage of NRS-M here is that individual scales can easily be added or removed without necessarily affecting overall mood score, so that it could readily be adapted as appropriate for use in different contexts – for example, by including further positive valence moods or feelings. Other ways in which NRS-M may be preferable to $BRUMS_{24}$ as a measure of mood – particularly its sensitivity – are summarised above and under the 'Conclusions' heading below. However, various scales or subscales may need to be carefully considered, and possibly piloted, before making a final selection for research use.

In our opinion, NRS-M is a flexible and sensitive tool that is appropriate for acupuncture and other research, in a variety of settings. Studies already exist relating HRV and mood, for example (Catipović-Veselica et al. 1999). Our own future plans include assessment of the external (criterion) validity of NRS-M in relation to HRV and other measures such as those extracted from EEG data, as well as of its test-retest reliability.

Conclusions

In this study, concurrent validity of NRS-M subscales was generally good when compared with the corresponding subscales of $BRUMS_{24}$. There was convergence between NRS-M 'Comfortable' and 'Relaxed', but otherwise NRS-M appeared to have greater discriminant validity than $BRUMS_{24}$. Internal consistency of NRS-M was of course less than for $BRUMS_{24}$, and low test-retest reliability suggests that NRS-M may be more sensitive to changes in mood than $BRUMS_{24}$. Various additional methods strongly indicated the greater sensitivity of NRS-M to mood and its changes, although significances of Pre-Post differences were very similar for both measures (albeit with a small effect size). Factor and cluster analysis of NRS-M were reasonably consistent, particularly for Pre-Post changes, clearly separating moods of positive and negative affect. Users (and younger men in particular) completed NRS-M in markedly less time than they did $BRUMS_{24}$.

Response style (RS) and problems such as reverse scoring were not thought to threaten the validity of these conclusions, although this warrants further study, as do associations between RS and emotional intelligence (as assessed using the TEIQue-SF, for example). The small qualitative vignette investigating in what way personal characteristics and attitudes to measures of mood may affect how such moods are scored could usefully be developed.

In summary, NRS-M is a reasonably robust, responsive and partially validated measure of mood to use in the context of acupuncture practice. It has advantages over longer Likert scales such as $BRUMS_{24}$.

Acknowledgments

To **Andrea Abele-Brehm** for providing an English version of her *Befindlichkeitsskale* (mood scale); to **Helmut Acker** for information about this scale; to **Matthew Hankins** for making available his Excel spreadsheet for calculating $delta_G$ from frequency tables, and for his advice on its use; to **Bev Steffert** for information on dyscalculia; to **Charles van Wijk** for clarification on 'total mood distress'; to **Steve Vodanovich** for his thoughts on boredom; to **Tim Watson** and of course our spouses/partners, and our questionnaire respondents, without whom none of this would have been possible. It goes without saying that any errors in this report remain our own responsibility.

References

Abele-Brehm A, Brehm W. 1986. Zur Konzeptualisierung und Messung von Befindlichkeit. Die Entwicklung der "Befindlichkeitsskalen" (BFS). Diagnostica, 32(3), 209–228.

Abele-Brehm A. n.d. (1) Items of BFS in German and in English; (2) BFS short form (English). Supplied by author, 23 Feb 2016.

Abend R, Dan O, Maoz K et al. 2014. Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. Journal of Behavior Therapy and Experimental Psychiatry, 45(4), 447-453.

Abu-Bader S. 2010. Advanced and Multivariate Statistical Methods for Social Science Research. Chicago, IL: Lyceum.

Acker H, Schmidt-Rathjens C, Acker T et al. 2015. Acupuncture-brain interactions as hypothesized by mood scale recordings. Medical Hypotheses, 85(3), 371-379.

Ahearn EP, Carroll BJ. 1996. Short-term variability of mood ratings in unipolar and bipolar depressed patients. Journal of Affective Disorders, 36(3-4), 107-15.

Alston W. 1967. Emotion and feeling. In Edwards P (Ed.). The Encyclopedia of Philosophy (Vol 2). New York: Macmillan.

Arruda JE, Stern RA, Somerville JA. 1999. Measurement of mood states in stroke patients: validation of the visual analog mood scales. Archives of Physical Medicine and Rehabilitation, 80(6), 676-680.

Bauer MS, Crits-Christoph P, Ball WA et al. 1991. Independent assessment of manic and depressive symptoms by self-rating. Scale characteristics and implications for the study of mania. Archives of General Psychiatry, 48(9), 807-812.

Baumgartner H, Steenkamp J-BEM. 2001. Response styles in marketing research: a cross-national investigation. Journal of Marketing Research, 38(2), 143-156.

Beedie CJ, Terry PC, Lane AM. 2005. Distinctions between emotion and mood. Cognition and Emotion, 19(6), 847-878.

Cabooter EFK. 2010. The impact of situational and dispositional variables on response styles with respect to attitude measures. Dissertation submitted to the Faculty of Economics and Business Administration, Ghent University, in fulfillment of the requirements for the degree of Doctor in Applied Economic Sciences.

Carlson CR, Collins Jr, FL, Stewart JF et al. 1989. The assessment of emotional reactivity: A scale development and validation study. Journal of Psychopathology and Behavioral Assessment, 11(4), 313-325.

Carpenter RJ, Dillard J, Zion AS et al. 2010. The acute effects of acupuncture upon autonomic balance in healthy subjects. American Journal of Chinese Medicine, 38(5), 839-847.

Catipović-Veselica K, Amidzić V, Durijancek J et al. 1999. Association of heart rate and heart-rate variability with scores on the emotion profile index in patients with acute coronary heart disease. Psychological Reports, 84(2), 433-442.

Cella DF, Perry SW. 1986. Reliability and concurrent validity of three Visual-Analogue Mood Scales. Psychological Reports, 59(2 pt 2), 827-833.

Cella DF, Jacobsen PB, Orav EJ et al. 1987. A brief POMS measure of distress for cancer patients. Journal of Chronic Diseases, 40(10), 939-942.

Crist DA, Rickard HC, Prentice-Dunn Set al. 1989. The Relaxation Inventory: self-report scales of relaxation training effects. Journal of Personality Assessment, 53(4), 716-726.

DeBruine L. n.d. Reporting Statistics in Psychology. Face Research Lab, Institute of Neuroscience and Psychology, University of Glasgow. Available online at: http://facelab.org/debruine/Teaching/Meth_A/files/Reporting_Statistics.pdf [accessed 29 Jan 2016].

DeVon HA, Block ME, Moyle-Wright P et al. 2007. A psychometric toolbox for testing validity and reliability. Journal of Nursing Scholarship, 39(2), 155-164.

Ferguson GA. 1949. On the theory of test discrimination. Psychometrika, 14(1), 61-68.

Hall DP, Benedek D, Chang A. 1996. Ultradian cycles of mood in normal and depressed subjects. Jefferson Journal of Psychiatry. Jefferson Journal of Psychiatry, 13(1), Article 3. http://jdc.jefferson.edu/jeffjpsychiatry/vol13/iss1/3 [Accessed 13 Feb 2016].

Hankins M. 2007. Questionnaire discrimination: (re)-introducing coefficient delta. BMC Medical Research Methodology, 7:19.

Hankins M. 2008. How discriminating are discriminative instruments? Health and Quality of Life Outcomes, 6:36.

Heponiemi T, Keltikangas-Järvinen L, Puttonen S et al. 2005. Vital exhaustion, temperament, and the circumplex model of affect during laboratory-induced stress. Cognition and Emotion, 19(6), 879-897.

Jacobi J. 1971. Complex / Archetype / Symbol in the Psychology of C.G. Jung. Princeton, NJ: Princeton University Press.

Kolcaba KY. 1991. A taxonomic structure for the concept comfort. ImageL The Journal of Nursing Scholarship, 23(4), 237-240.

Kolcaba K, Steiner R. 2000. Empirical evidence for the nature of holistic comfort. Journal of Holistic Nursing, 18(1), 46-62.

Lagae L. 2008. Learning disabilities: definitions, epidemiology, diagnosis, and intervention strategies. Pediatric Clinics of North America, 55(6), 1259-1268.

Lane AM, Jackson A, Terry PC. 2005. Preferred modality influences on exercise-induced mood changes. Journal of Sports Science and Medicine, 4(2), 195-200.

Li QQ, Shi GX, Xu Q et al. 2013. Acupuncture effect and central autonomic regulation. Evidence-Based Complementary and Alternative Medicine (eCAM). 2013;2013:267959.

Lorr M. 1989. Models and methods for measurement of mood. In: Plutchik R, Kellerman H (Eds.). The Measurement of Emotion. San Diego, CA: Academic Press, 37-53.

Luutonen S. 2007. Anger and depression – theoretical and clinical considerations. Nordic Journal of Psychiatry, 61(4), 246-251.

McNair DM, Lorr M, Droppleman LF. 1971. Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service.

McNair DM, Lorr M, Droppleman LF. 1992. Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service.

Maciocia G. 2009. The Psyche in Chinese Medicine. Treatment of emotional and mental disharmonies with acupuncture and Chinese herbs. Edinburgh: Churchill Livingstone.

Mayer JD, Gaschke YN. 1988. The experience and meta-experience of mood. Journal of Personality and Social Psychology, 55(1), 102-111.

Mayor DF. 2007. (Ed.). Electroacupuncture. A practical manual and resource. Edinburgh: Churchill Livingstone, 66.

Mayor DF. 2014. Expectation and experience of 'nonspecific' (whole person) feelings elicited by acupuncture: Content validity of a set of questionnaires. Deutsche Zeitschrift für Akupunktur/ German Journal of Acupuncture and Related Techniques, 57(1), 14-19.

Mayor D, Steffert T. 2013. Expectation and experience of the 'nonspecific' effects of acupuncture: developing and piloting a set of questionnaires. Poster, BMAS/PMAS Meeting, Porto. http://www.qeeg.co.uk/electroacupuncture/eaq1.htm [accessed 12 March 2016].

Mayor D, Steffert T, Bhavsar R. 2015. Changes in finger temperature and blood flow in response to different frequencies of transcutaneous electroacupuncture at LI4 (hegu). Interim analysis and 'real life' methodological issues: many factors, missing data and a multiplicity of measures. Poster, ARRC Symposium, London. http://www.qeeg.co.uk/electroacupuncture/bloodflow.htm [accessed 12 March 2016].

Mayor DF, McClure LS. Nonspecific feelings expected and experienced during electroacupuncture. A pilot study in a teaching situation of questionnaires revised following an exercise to establish content validity. (In preparation.)

Monchi R, Suzuki N. 2000. [The factor structure of relief feeling in situations: relief from tension and calmness and relaxation]. Shinrigaku Kenkyu, 71(1), 42-50.

Norman GR. 2008. Discrimination and reliability: equal partners? Health and Quality of Life Outcomes, 6:81.

Nowlis V. 1970. Mood behavior and experience. In: Arnold MB (Ed.). Feelings and Emotions (The Loyola Symposium). New York: Academic Press, 261-278.

Okun MS, Green J, Saben R et al. 2003. Mood changes with deep brain stimulation of STN and GPi: results of a pilot study. Journal of Neurology, Neurosurgery, and Psychiatry, 74(11), 1584-1586.

Petrides KV. 2009. Psychometric properties of the Trait Emotional Intelligence Questionnaire (TEIQue). In: Stough C, Saklofske DH, Parker JDA. (Eds.). Assessing Emotional Intelligence. Theory, research, and applications. New York: Springer Science+Business Media, New York, Philadelphia, 85-101.

Plutchik R. 1958. Outlines of a new theory of emotions. Psychosomatic Medicine, 17(**issue**), 306-310.

Plutchik R, Platman SR, Fieve RR. 1968. Repeated measurements in the manic depressive illness: some methodological problems. Journal of Psychology, 70(1st half), 131-137.

Plutchik R, Kellerman H. 1984. Emotions Profile Index Manual. Los Angeles, CA: Western Psychological Services.

Plutchik R. 1989. Measuring emotions and their derivatives. In: Plutchik R, Kellerman H (Eds.). The Measurement of Emotion. San Diego, CA: Academic Press, 1-35.

Posner J, Russell JA, Peterson BS. 2005. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17(3), 715-34.

Reigosa-Crespo V, Valdés-Sosa M, Butterworth B et al. 2012. Basic numerical capacities and prevalence of developmental dyscalculia: the Havana Survey. Developmental Psychology, 48(1), 123-135

Rossi V, Pourtois G. 2012. Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: a comparative review. Anxiety, Stress, and Coping, 25(6), 603-645.

Russell JA. 1980. A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178.

Sakakibara M, Teramoto Y, Tani I. 2014. [Development of a short-form self-report measure to assess relaxation effects]. Shinrigaku Kenkyu, 85(3), 284-293.

Salinsky MC, Storzbach D, Dodrill CB et al. 2001. Test-retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12-16-week period. Journal of the International Neuropsychological Society, 7(5), 597-605.

Sawazaki K, Mukaino Y, Kinoshita F et al. 2008. Acupuncture can reduce perceived pain, mood disturbances and medical expenses related to low back pain among factory employees. Industrial Health, (4), 336-340.

Shacham S. 1983. A shortened version of the Profile of Mood States. Journal of Personality Assessment, 47(3), 305-306.

Shalev RS, Gross-Tsur V. 2001. Developmental dyscalculia. Pediatric Neurology, 24(5), 337-342.

Schilling DJ, Poppen R. 1983. Behavioral relaxation training and assessment. Journal of Behavior Therapy and Experimental Psychiatry, 14(2), 99-107.

Schlosberg H. 1952. The description of facial expressions in terms of two dimensions. Journal of Experimental Psychology, 4(4), 229-237.

Siegel LS. 2006. Perspectives on dyslexia. Paediatrics and Child Health, 11(9), 581-587.

Smith JC, Wedell AB, Kolotylo CJ et al. 2000. ABC Relaxation Theory and the factor structure of relaxation states recalled relaxation activities, dispositions, and motivations. Psychological Reports. 86(3 pt 2), 1201-1208.

Smith JC. 2001. Advances in ABC Relaxation: Applications and inventories. New York: Springer.

Steffert T, Holland S, Mulholland P, Väljamäe A. n.d. The assessment of real-time EEG sonification for neurofeedback. (Submitted for publication.)

Steffert T, Mayor D. The fickleness of data: Estimating the effects of different aspects of acupuncture treatment on heart rate variability (HRV). Initial findings from three pilot studies. Poster, ARRC

Symposium, London. http://www.qeeg.co.uk/electroacupuncture/hrv1.htm [accessed 12 March 2016].

Stem RA. 1997. Visual Analog Mood Scales. Odessa,FL: Psychological Assessment Resources, Inc.

Terry PC, Lane AM, Lane HJ et al. 1999. Development and validation of a mood measure for adolescents. Journal of Sports Sciences, 17(11), 861-872.

Terry PC, Lane AM, Fogarty GJ. 2003. Construct validity of the Profile of Mood States — Adolescents for use with adults. Psychology of Sport and Exercise, 4 (2), 125–139.

van Rosmalen J, van Herk H, Groenen PJF. 2010. Identifying response styles: a latent-class bilinear multinomial logit model. Journal of Marketing Research, 47(1), 157-172.

Van Vaerenbergh Y, Thomas TD. 2013. Response styles in survey research: a literature review of antecedents, consequences, and remedies. International Journal of Public Opinion Research, 25(2), 195-217.

Van Wijk C. 2011. The Brunel Mood Scale: A South African norm study. South African Journal of Psychiatry, 17(2), 44-54.

Van't Noordende JE, van Hoogmoed AH, Schot WD et al. 2015. Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. Psychological Research [Epub ahead of print]

Vodanovich SJ, Watt JD. 2016. Self-report measures of boredom: an updated review of the literature. Journal of Psychology, 150(2), 194-226.

Watson D, Clark LA, Tellegen A. 1988. Development and validation of brief measures of Positive and Negative Affect: The PANAS Scales. Journal of Personality and Social Psychology, 54(6), 1063-1070.

Watson D, Clark LA. 1999. THE PANAS-X. Manual for the Positive and Negative Affect Schedule - Expanded Form. University of Iowa. http://www2.psychology.uiowa.edu/Faculty/Clark/PANAS-X.pdf [accessed 12 March 2016]

Weijters , Schillewaert N, Geuens M. 2008. Assessing response styles across modes of data collection. Journal of the Academy of Marketing Science, 36(3), 409-422.

Weijters , Geuens M, Schillewaert N. 2010a. The stability of individual response styles. Psychological Methods, 15(1), 96–110.

Weijters B, Cabooter E, Schillewaert N. 2010b. The effect of rating scale format on response styles: The number of response categories and response category labels. International Journal of Research in Marketing, 27(3), 236–247.

Wetzel E, Böhnke JR, Carstensen CH, Ziegler M, Fritz Ostendorf F. 2013. Do individual response styles matter? Assessing differential item functioning for men and women in the NEO-PI-R. Journal of Individual Differences , 34(2), 69-81.

Whitmont EC. 1978. The Symbolic Quest. Basic concepts of analytical psychology. Princeton, NJ: Princeton University Press.

Williams VS, Morlock RJ, Feltner D. 2010. Psychometric evaluation of a visual analog scale for the assessment of anxiety. Health and Quality of Life Outcomes, 8:57.

Wyrwich KW. 2008. Understanding the role of discriminative instruments in HRQoL research: can Ferguson's Delta help? Health and Quality of Life Outcomes, 6:82.

Zelenbrz J. 2005. Provjera Plutchikovog Indeksa Profila Emocija)IE) u selekcijskoj situaciji [Testing the Emotion Profile Index (EPI) in the selection context]. Diploma Thesis, Department of Psychology, Filozofski fakultet u Zagrebu.

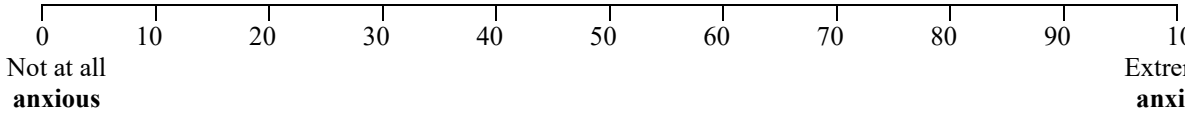
Zhang CQ, Si G, Chung PK et al. 2014. Psychometric properties of the Brunel Mood Scale in Chinese adolescents and adults. Journal of Sports Sciences, 32(15), 1465-1476.

Appendices

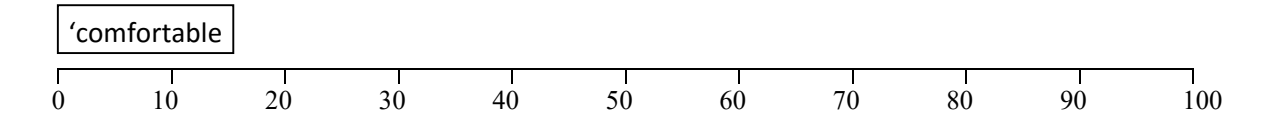
Appendix A. Multiple numerical rating scale for mood (NRS-M)

Place a cross on each line to represent **how you feel right now**.

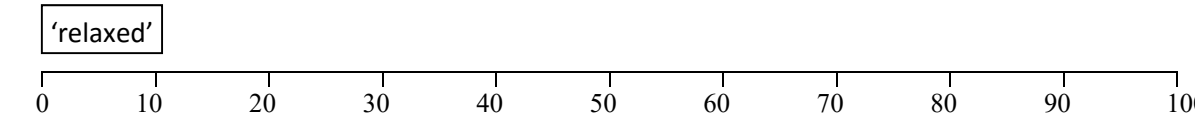
‘anxious’



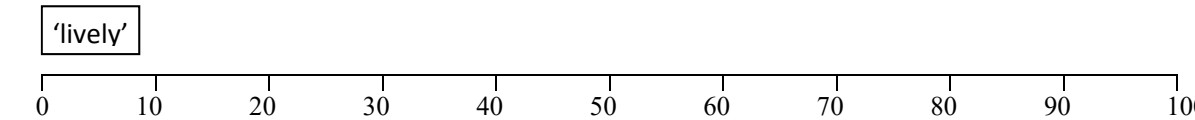
‘comfortable



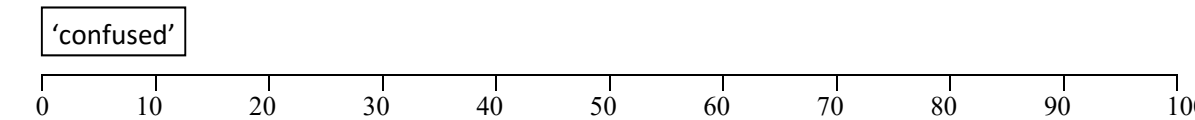
‘relaxed’



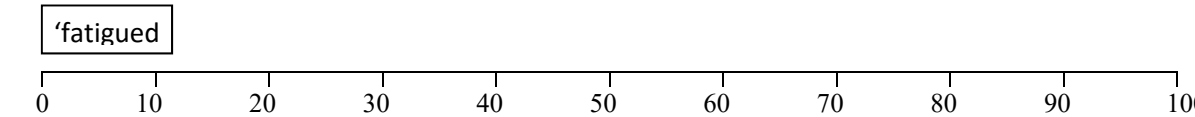
‘lively’



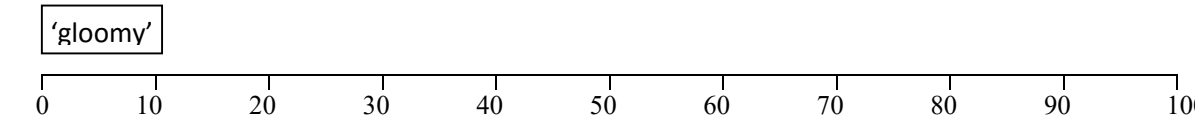
‘confused’



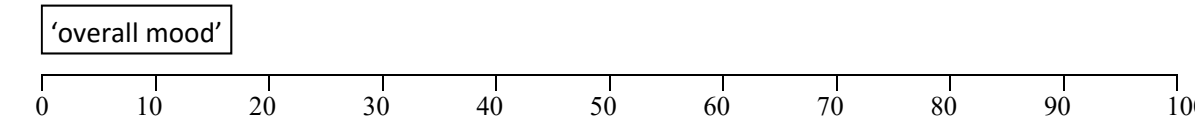
‘fatigued



‘gloomy’



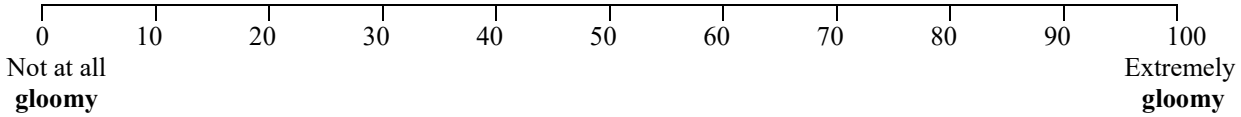
‘overall mood’



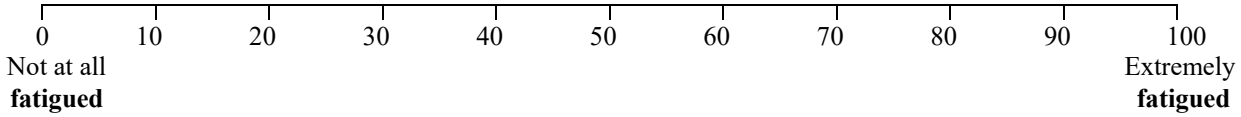
ID _____ Date _____

Place a cross on each line to represent **how you feel right now**.

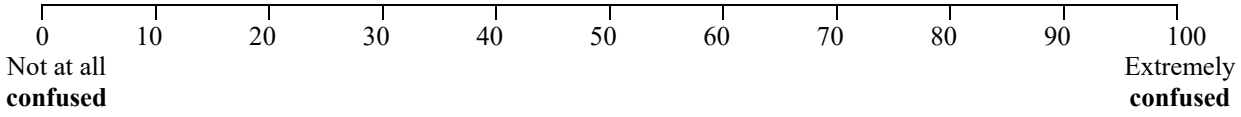
'gloomy'



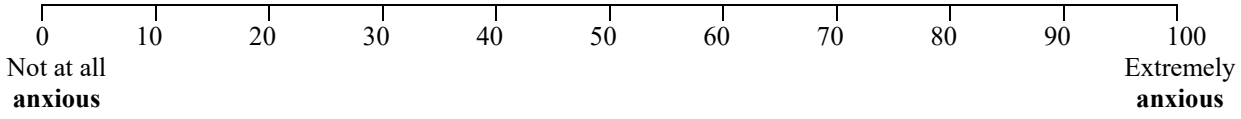
'fatigued



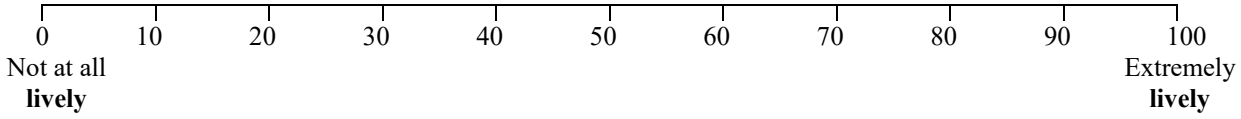
'confused'



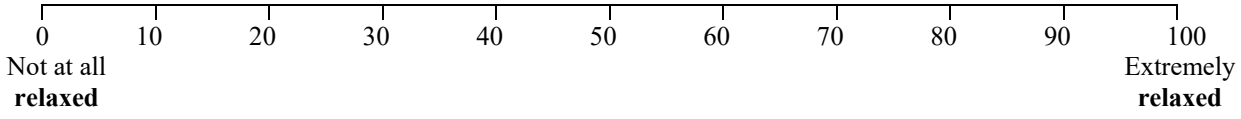
'anxious'



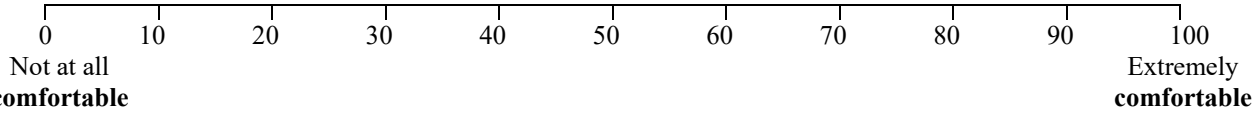
'lively'



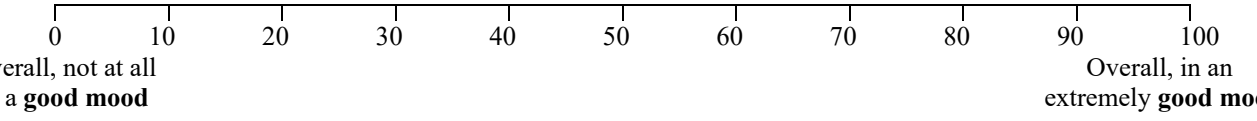
'relaxed'



'comfortable



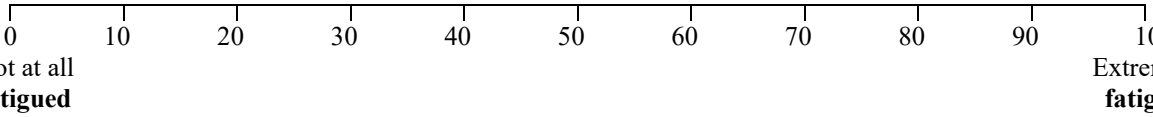
'overall mood'



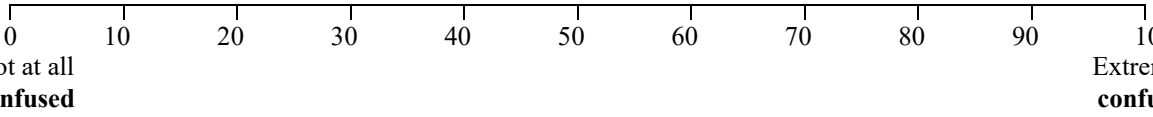
ID _____ Date _____

Place a cross on each line to represent **how you feel right now**.

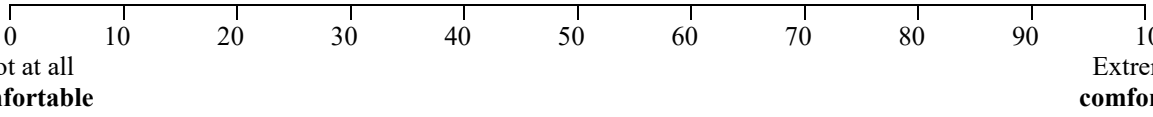
'fatigued



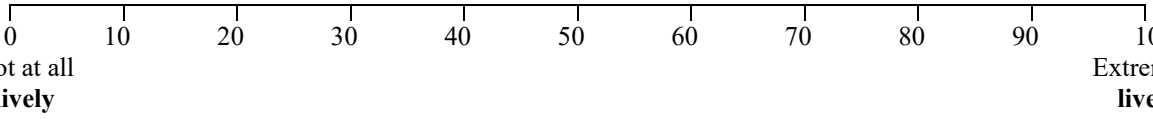
'confused'



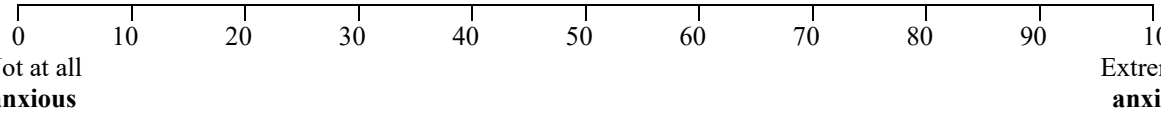
'comfortable



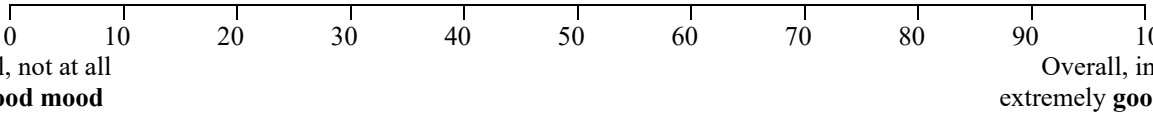
'lively'



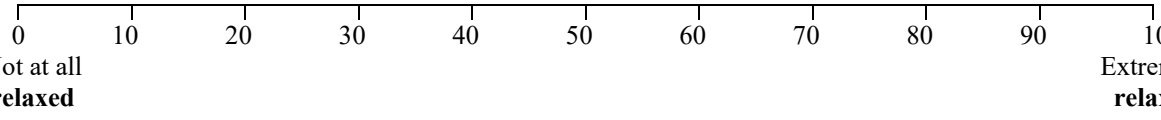
'anxious'



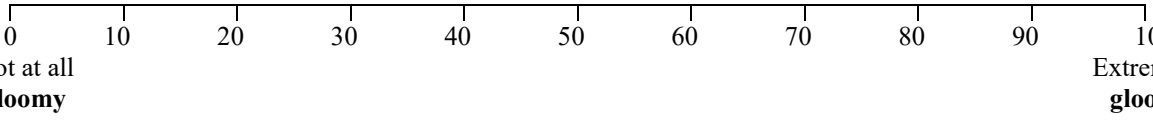
'overall mood'



'relaxed'



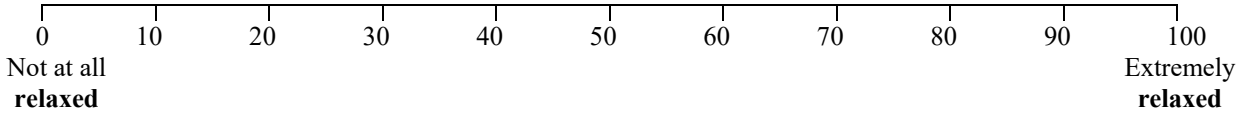
'gloomy'



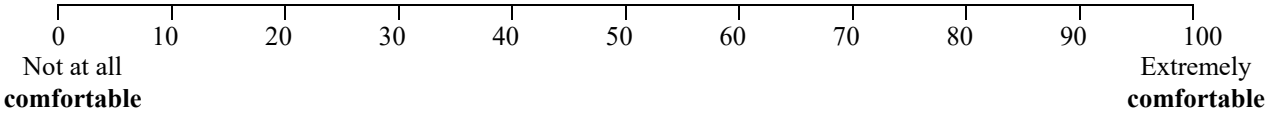
ID _____ Date _____

Place a cross on each line to represent **how you feel right now**.

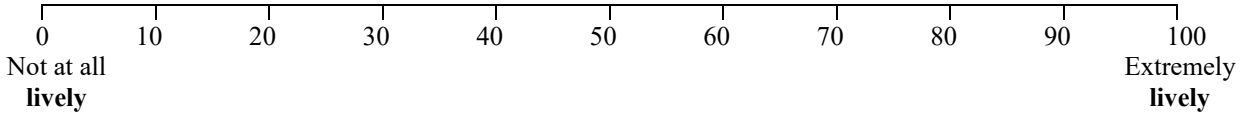
'relaxed'



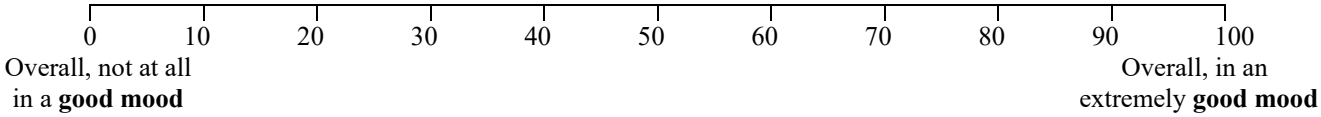
'comfortable



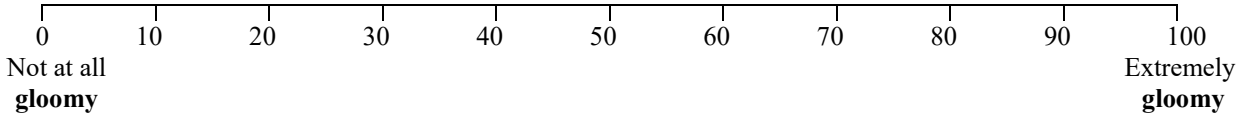
'lively'



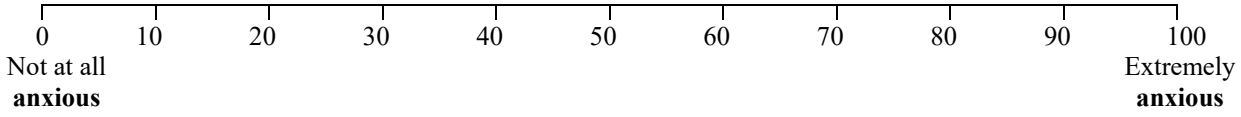
'overall mood'



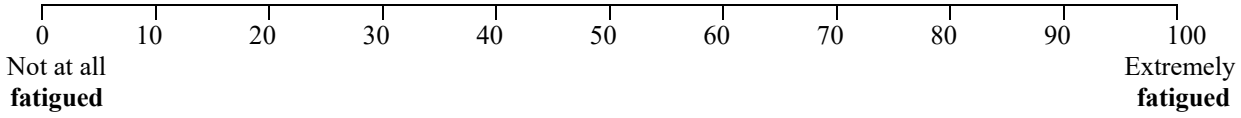
'gloomy'



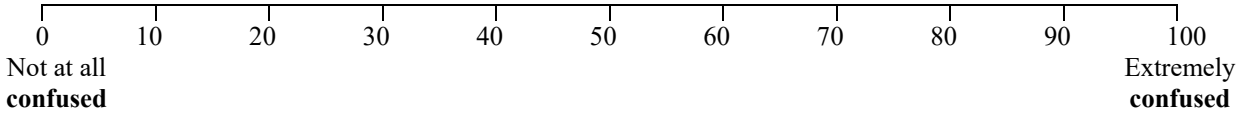
'anxious'



'fatigued



'confused'



ID _____ Date _____

Appendix B. Brunel Mood Scale ($BRUMS_{24}$)

Brunel Mood Scale ($BRUMS_{24}$)

Completing this questionnaire should take you less than 5 minutes.

Instructions

Below is a list of 24 words describing feelings. Please read each word carefully, and circle one of the numbers to the right of the word to describe **how you feel right now**.

Use this scale to record your answers:

Word	Not at all	A little	Moderately	Quite a bit	Extremely
e.g. Energetic	0	1	(2)	3	4

Please make sure that you circle a number for *every* word, and be as honest and accurate as you can throughout. Try not to let your response to one statement influence your responses to other statements. There are no 'correct' or 'incorrect' answers. Answer according to your own feelings, rather than how you think 'most people' would answer.

Word	Not at all	A little	Moderately	Quite a bit	Extremely
1. Panicky	0	1	2	3	4
2. Lively	0	1	2	3	4
3. Confused	0	1	2	3	4
4. Worn out	0	1	2	3	4
5. Depressed	0	1	2	3	4
6. Downhearted	0	1	2	3	4
7. Annoyed	0	1	2	3	4
8. Exhausted	0	1	2	3	4
9. Mixed-up	0	1	2	3	4
10. Sleepy	0	1	2	3	4
11. Bitter	0	1	2	3	4
12. Unhappy	0	1	2	3	4
13. Anxious	0	1	2	3	4
14. Worried	0	1	2	3	4
15. Energetic	0	1	2	3	4
16. Miserable	0	1	2	3	4
17. Muddled	0	1	2	3	4
18. Nervous	0	1	2	3	4
19. Angry	0	1	2	3	4
20. Active	0	1	2	3	4
21. Tired	0	1	2	3	4
22. Bad tempered	0	1	2	3	4
23. Alert	0	1	2	3	4
24. Uncertain	0	1	2	3	4

Date _____ **ID** _____

Thank you for completing this questionnaire.

Derivation of scores for $BRUMS_{24}$ subscales and Total Mood Distress (TMD) from those of the individual items:

A. ANGer	A = 7 + 11 + 19 + 22
C. CONFusion	C = 3 + 9 + 17 +24
D. DEPRession	D = 5 + 6 + 12 + 16
F. FATigue	F = 4 + 8 + 10 + 21
T. TENSion	T = 1 + 13 + 14 + 18
V. VIGour	V = 2 + 15 + 20 + 23
TMD	TMD = A + C +D + F + T – V