Expectation and experience of 'nonspecific' feelings elicited by acupuncture: Developing and piloting a set of questionnaires

© David Mayor (University of Hertfordshire) and Tony Steffert (Open University)

Abstract
32-item 'EXPectation of feelings' questionnaires were developed to explore the expected (EXPre) and reported (EXPost) incidence of 'nonspecific' (whole person) feelings in response to acupuncture-type interventions, in particular electroacupuncture (EA) and transcutaneous electrical acupoint stimulation (TEAS). They were tested on 204 participants familiar with acupuncture in three separate cohorts (Pilot, CPD, Students). Their validity and reliability were assessed, a cluster analysis conducted, and the association between expectation and experience analysed, along with those items most frequently found, association with other trait and state measures, and the influence of various aspects of treatments on reported feelings. Results are reported and discussed. Methods and further material are provided online at http://www.qeeg.co.uk/electroacupuncture/.

Introduction

Nonspecific effects are usually defined from the practitioners' point of view, as those associated with the incidental elements of an intervention (e.g. the overall therapeutic context rather than acupuncture needling *per se*), and are thus considered as synonymous with placebo effects (Linde et al., 2010). The placebo effect has been described as evidence for our self-healing capacities (Peters, 2001; Walach & Jonas, 2004). Expectation of positive outcome is often thought to be a major contributor to this nonspecific effect (Pacheco-López et al., 2006), in part because it alters how bodily sensations are identified (Geers et al., 2011). In response to placebo acupuncture, bodily sensations of warmth, tingling or flow have been elicited (Kerr et al., 2011). Such sensations have also been interpreted as resulting from the flow of *Qi* (Mayor, 2011).

From the acupuncture recipients' point of view, nonspecific effects may be those considered as those incidental to their desired treatment outcome. 'EXPectation of feelings' questionnaires were developed to explore the expected and reported incidence of such relatively nonspecific feelings (whether bodily, emotional or mental) in response to acupuncture-type interventions, irrespective of the model used to explain their occurrence, and whether or not they are considered to result from the essential or incidental ('placebo') elements of such interventions.

There do not appear to any existing questionnaires that can be used for this purpose. There is minimal overlap, for example, between the EXP questionnaires and scales designed to assess the specific sensations of acupuncture needling such as the Acupuncture Sensations Scale (Park et al., 2002; Park et al., 2005 [329]; Vincent et al., 1989) and the related Acupuncture Sensation Questionnaire (Kim et al., 2008), Southampton Needle Sensation Questionnaire (Pach et al., 2011; White et al., 2008), Subjective Acupuncture Sensation Scale (Kong et al., 2005) and Massachusetts General Hospital Acupuncture Sensation Scale (Kong et al., 2007; Yu et al., 2012). There is greater overlap with the recently published Treatment Experience Questionnaire (5 Mental state items, 3 Bodily sensation items) (Blasche et al., 2013), but this was developed for quite a different purpose.

Aim

To develop and pilot questionnaires designed to assess expectation and subsequent experience of nonspecific (whole person) feelings in response to electroacupuncture (EA) and transcutaneous electrical acupoint stimulation (TEAS).

Objectives

To select items appropriate for these questionnaires. To assess questionnaire validity and reliability. To examine possible clusters of items. To determine whether there is any association between expectation and experience, which items are expected or experienced most/least frequently, and their possible association with other trait and state measures. To test whether particular aspects of treatment influence the experienced feelings reported.

Methods

Two questionnaires, 'EXPre' and 'EXPost', were constructed on the basis of reports in the literature (e.g. Johnson 1973; Lindsay et al. 1984; Mayor, 2011), undergraduate projects (Morris, 2007; Vearncombe, 2007), and items extracted from standard (unrelated) questionnaires such as the Profile of Mood States (POMS) and SF-36 Health Survey. They had also been discussed informally with a small panel of qualitative researchers and in a brief focus group (8 Aug 2011, *N*=5). They included some items as distractors, and some which overlap in meaning. Both questionnaires consisted of the same 32 items (**Appendix I**), each referring to a feeling (e.g. peacefulness) or

sensation (e.g. tingling) either *expected* (EXPre) or *experienced* during or immediately after EA/TEAS stimulation (EXPost). In EXPre, respondents were asked whether they 'expect to experience a change in the feeling of …', and in EXPost, whether they 'experienced a change in the feeling of …'. Responses were limited to 'Yes' (Y), 'No' (N) or 'Don't know' (DK), but in EXPost they were also asked to asterisk 'those changes ... noticed most' (*). (Note: respondents were not being asked whether they expected or experienced an increase, decrease or improvement in feelings).

Participants

The questionnaires were tested in three different settings: (1) during an experimental study on the effects of EA/TEAS on the electrical activity of the brain and heart, in which participants were themselves acupuncture or other complementary health practitioners (*N*=21); (2) in seven course seminars for acupuncture students (*N*=129); (3) during four 'continuing professional development' (CPD) training days for acupuncture practitioners (*N*=54). In (1), stimulation characteristics (electrical parameters and acupoints) were standardised, but in (2) and (3) were selected by participants, although under guidance.

Our experimental study was divided into four Pilots: Pilot 1 (*N*=8), in which five participants attended for two sessions of TEAS around one week apart (one at 2.5 Hz, the other at 10 Hz) and two for one session only (one at 2.5 Hz, the other at 10 Hz), with one participant completing an online version of EXPre but not attending further; Pilot 2 (*N*=12), in which all participants attended for four sessions incorporating both manual acupuncture (MA) and EA (two at 2.5 Hz, two at 10 Hz), 1-7 weeks apart; Pilot 3 (*N*=4), in which participants from Pilot 1 were re-invited to attend for four sessions of EA and TEAS (two at 2.5 Hz, two at 10 Hz), again 1-7 weeks apart; and Pilot 4, in which one participant attended for six sessions of auricular TEAS, at irregular intervals. The acupoints used in Pilots 1-3 were LI4 and ST36, in various combinations.

Validity

Given the evanescence and subjectivity of 'feelings', it is difficult if not impossible to establish the validity of a scale to assess their expectation or experience. In particular, no existing measure of 'expectation of feelings' could be found, so criterion-related validity could not be established. In addition, because these questionnaires were not designed to capture a specific construct (other than 'expectation/experience of change'), their construct validity could not be assessed either. However, with the confidence of two years' experience with the questionnaires, a survey was conducted to assess **content validity**, following the method first described by Lawshe (1975).

Twenty experienced acupuncture practitioners or researchers were invited to rate 48 items (the original 32 together with 16 other possible candidates, listed in **Appendix II**) as 'essential', 'useful but not essential', or 'not necessary' for inclusion in a list of the nonspecific effects of acupuncture (the term 'nonspecific' was not defined). They were given the opportunity to add further items if desired, and asked nine other brief questions about themselves and potential applications for the questionnaires.

A content validity ratio (CVR) and content validity index (CVI_L) were calculated using Lawshe's method, with the correction by Wilson et al. (2012), and also a CVI using Lynn's method (Lynn, 1986), scoring the same responses as a dichotomy, but with 'useful but not essential' considered alternatively as either Essential or Inessential. This resulted in an item CVI (I-CVI), and a score CVI (S-CVI), the latter calculated in three different ways. Lawshe's and L:ynn's versions of CVI – each of which has been used in prior acupuncture-related studies (Yu et al., 2012; Kim et al., 2008) – were then compared.

Reliability

Various methods of assessing reliability were used:

(1) **Inter-rater reliability** was assessed using Justus Randolph's free-marginal multi-rater *kappa,* κ_{free}, for non-summated categorical data (Randolph, 2008), on the basis that participants are scoring data from the same entity ('expected change' or 'experienced change') when using the questionnaires. This method of assessing *kappa* also assumes that all categories are equally likely (Warrens, 2010), which is in principle true of the questionnaire data. 'Free-marginal' rather than 'fixed-marginal' *kappa* was used because the number of items for each response category

(Y, N, DK, *) was not fixed (Brennan & Prediger, 1981). Values of $\kappa_{free} > 0.7$ are taken to indicate good inter-rater reliability (with values > 0 as better than chance).

Differences between κ_{free} for EXPre and EXPost in the different cohorts were tested for significance using the Wilcoxon test for related samples (confirmed with the Mann-Whitney U test when EXPre/EXPost cohorts were not completely identical). κ_{free} was calculated for individual items, and for all items taken together (computationally equivalent to the mean of the individual κ_{free} values). Lists were made of those items with the five highest and five lowest κ_{free} values for each cohort ('tied' items which would make these counts unwieldy were omitted).

(2) **Test-retest reliability**. It was possible to calculate this for EXPre only for the four participants who attended for Pilot 1 and then also Pilot 3, a year later. EXPost, on the other hand, was used after all visits by all participants in Pilots 1-4, so both short- and long-term reliability were examined. Results were obtained for both Spearman's *rho* (ρ) and Kendall's *tau*-b (τ).

(3) **Split-half reliability**. To assess whether respondents began to flag and lose attention when having to complete these quite long questionnaires, split-half reliability was computed for the 32 items, split in two different ways: (a) alphabetically by name; and (b) randomly.

(4) **Internal consistency reliability** (Cronbach's *alpha*). *Alpha* is widely accepted as a measurement of the internal consistency (reliability) of a multivariate measure in which several items are highly inter-correlated (DeVellis, 1991). A value of 0.8 for *alpha* is usually considered the criterion for internal reliability (Bryman & Cramer, 2001), although an *alpha* that is ≥0.7 is 'acceptable' (George and Mallery, 2003). A sample size of several hundred is generally considered necessary for *alpha* to be stable (Nunnally & Bernstein, Yurdugül, 2008), and furthermore *alpha* may be artificially inflated for a questionnaire containing more than a few items (Borgatti, 2008). Despite these drawbacks, it may have some value for comparative purposes as an index of internal consistency even for quite small samples.

Cluster analysis

The EXPre and EXPost questionnaire data is not amenable to factor analysis (categorical, not normally distributed, large number of items, insufficiently large sample), so exploratory and confirmatory cluster analyses were undertaken partway through data gathering to assess feasibility of this alternative approach. Further exploratory cluster analyses were undertaken for the full dataset. After some initial experimentation, an agglomerative hierarchical fourfold clustering method was used for EXPre and EXPost (all cases) based on Ward's method, with a chi-squared measure for categorical (count) data (Everitt, 1980). For the exploratory analysis, allowing a range of solutions (3-10 clusters) rather than forcing just one (4 clusters only) was found to result in greater stability of the clusters when item order was shuffled.

Cluster allocation was simple when it was consistent for an item over all three cohorts and the whole sample (e.g. 'Aliveness' was allocated to cluster 1 for all four participant groupings). If it was consistent for three of these (e.g. 1, 1, 1, 3 for two cohorts and the whole sample), this was taken as the basis for allocation. If it was consistent for only two, with the other two cohorts showing different allocations (e.g. 1, 1, 2, 4), this lesser agreement was taken as the basis for allocation. If, however, consistency was split evenly among the four groupings (e.g. 2, 2, 3, 3), then the item was allocated to two alternative clusters, here either cluster 2 or cluster 3.

In addition, a less formal attempt was made to extract clusters manually from the EXPre Pilot cohort data on the basis of positive inter-item correlations (IICs) ≥0.5 and counts of negative IICs.

The following clusters were adopted for confirmatory analysis:

Group A. Polarity style

[1] Items which might be construed as 'negative' in some way
[2] Items which might be construed as 'positive' in some way
[3] Items which might be construed as neither 'negative' nor 'positive'.

Group B. Feeling style

[4] 'bodily' feelings
[5] 'emotional' feelings
[6] 'mental' feelings
[7] 'general' feelings (interpretable as any of [4] to [6])

[8] A further cluster was also created, for items considered to relate to the construct 'relaxation' .

These clusters are detailed in **Appendix III** at the end of this document.

Cronbach's *alpha* and IIC were calculated for all clusters (averaged when inconsistencies between cohorts resulted in several possible cluster allocations, as described above).

Association between expectation and experience

Counts were made of those items for which Y, N or DK scores were found in EXPre and EXPost. The different combinations (EXPre→EXPost 'Yes'→'Yes', 'No'→'Yes', etc.) were coded as shown in **Table 1**, and the results tabulated and expressed graphically.

Table 1. EXPre→EXPost response codes (difference scores).

	EXPre Y (8)	EXPre N (1)	EXPre DK (10)
EXPost Y (7)	1	–6	3
EXPost N (2)	6	–1	8
EXPost DK (15)	–7	–14	–5

Most/least frequently found items

Counts of the five most and least frequently checked items were conducted (rather than those in the highest and lowest quartiles or deciles). If there were tied counts, up to nine items in total were included under 'most' or least', but if tied items would increase the total beyond this, these were not listed. Items were included if most or least frequently checked in the whole sample, two or three of the subsample cohorts (Pilots, Students, CPD), or in one Pilot and either the Student or CPD cohort. A chi-square analysis was conducted to test difference from an expected distribution of 1/3 each for 'Yes', 'No' and 'Don't know' responses.

Association with other trait and state measures

For Pilot participants, most frequently found items (and some cluster scores), along with total numbers of Y and N responses, were compared with scores from other completed scales to assess possible associations between them. The additional scales completed were:

Trait

BFI	The Big Five Inventory (Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness) (John et al., 2011)
BIS/BAS	Behavioural Inhibition/Behavioural Approach Systems (Drive, Fun seeking, Reward responsiveness) (Carver et al., 1994)
LOT-R	Life Orientation Test (Optimism) (Scheier et al, 1994)
MISS	Multidimensional Iowa Suggestibility Scale (3 subscales: Persuadability, Sensation contagion, Physiological reactivity; 2 companion scales: Psychosomatic control, Stubborn opinionatedness) (Kotov et al., 2004)

State

PSS-10	Perceived Stress Scale (during the *past month*) (Cohen et al., 1988)
POMS-SF	Profile of Moods State (Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigour-Activity, Fatigue-Inertia, Confusion-Bewilderment, Total Mood Disturbance, before and after stimulation) (Shacham, 1983)
VAS-R	Visual analogue scale for Relaxation (over *past month*, and before and after stimulation)

Influence of treatment aspects on reported feelings

Which feelings are expected or experienced could depend on a variety of factors, such as personality, or different aspects of the treatment itself, such as the location, frequency and amplitude of stimulation. A preliminary analysis was conducted to see if this was the case, checking for correlations between EXPre scores and participant ID, and between EXPost scores (and numbers of asterisked items) and visit number, stimulation frequency or location. Goodman and Kruskal's *lambda* coefficient was used as a measure of association (*phi* was also tried initially, but often appeared significant when *lambda* was not). To assess whether numbers of 'Yes' scores, for example, depended on treatment aspects (i.e. an assessment of association between interval and ordinal data), the correlation ratio *eta* was used. To check whether numbers of items asterisked differed significantly between visits or for the two stimulation frequencies, because different numbers of participants attended for visits in Pilot 1 the independent samples Kruskal-Wallis and Mann-Whitney U tests were used rather than their matched-pair equivalents.

SPSS v.20 and Excel v.14 software were used to generate statistics.

Results

Validity: Content validity

The five items most frequently scored as 'essential' were 'Pain'*, 'Relaxation'*, 'Relief', 'Calmness'* and 'Tension' (those asterisked also showed high survey inter-rater reliability, κ_{free}). Lawshe's CVR was only significant for the first two of these, and then CVI_L only approached an acceptable value if those scoring no item as 'essential' were excluded. Results for CVR and CVI_L were better for women than for men respondents. Similar results were obtained for Lynn's I-CVI and S-CVI when 'useful' responses were scored as Inessential. However, when 'useful' responses were scored as Essential, more items showed acceptable values of I-CVI. Even so, '$S\text{-}CVI_{AV\text{-}UA}$', a measure combining the standard 'universal agreement' (UA) and 'averaging' (AV) methods of calculating S-CVI (Polit et al., 2007), was greater than the recommended 0.9 benchmark only for women respondents (for 30 items for all women respondents, and for 34 items for women practitioners without researchers). $S\text{-}CVI_{AV\text{-}UA}$ approached 0.9 for all respondents taken together and for men only, particularly if researchers were excluded from analysis (9 items and 7 items, respectively). Results of the content validity survey will be published elsewhere.

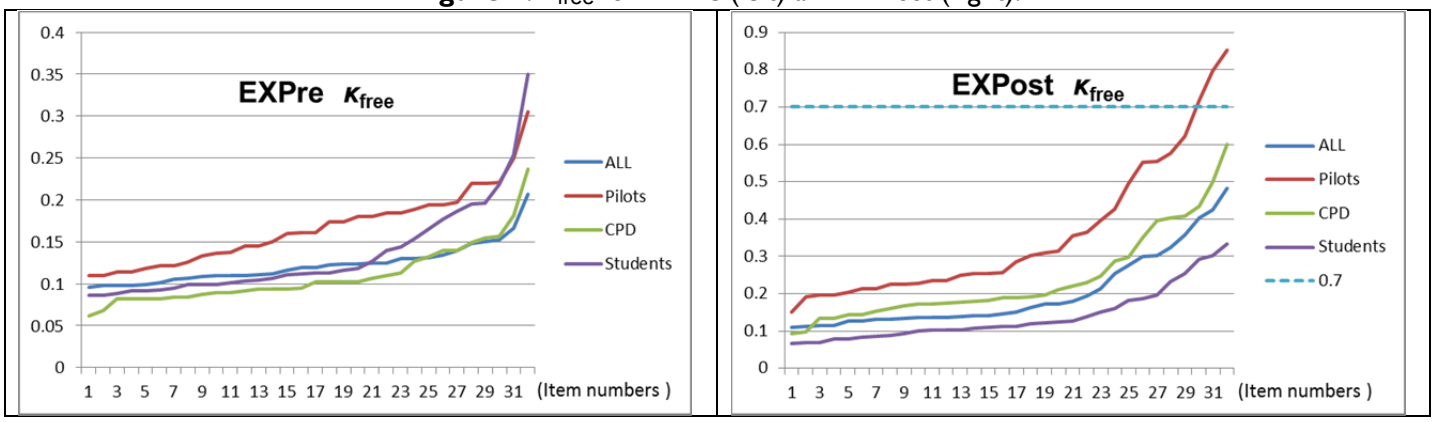
Reliability: Inter-rater reliability

Given the subjective nature of these questionnaires, good inter-rater reliability for all items taken together was not expected, and in fact was low (though positive) for all EXPre cohorts (range 0.06-0.14, 0.12 for the whole sample), but somewhat higher for the EXPost cohorts (range 0.14-0.36, 0.20 for the whole sample; all EXPre/EXPost comparisons significant except for the Student cohort). Greatest EXPre→EXPost increase was found for the Pilots cohort (472%, $p<0.001$), and least for the Students (3.5%).

Thus, as might be expected, there was a less consistent change in inter-rater agreement between EXPre and EXPost among acupuncture students than among experienced practitioners (even though some of those in Pilot 2 were not practitioners of acupuncture, but of other complementary therapies).

Calculation of κ_{free} for the individual items showed (as for all items taken together) that this was less for EXPre than for EXPost, but also that it only reached significance (>0.7) for EXPost in the Pilots and not the other cohorts. The charts below illustrate that there was a longer 'tail' of low κ_{free} in the CPD groups than the others for EXPre, and a longer 'tail' in the Student groups for EXPost, suggesting different degrees of inter-rater reliability in the different cohorts, with low κ_{free} for more items among students and those attending CPD seminars than those taking part in the Pilots.

Figure 1. κ_{free} for EXPre (left) and EXPost (right).



In all cohorts, the 'top five' items with the highest κ_{free} were mainly those items scored N (EXPre or EXPost) or DK (EXPre). Only a few EXPre Y items showed high κ_{free} (and no EXPost Y items), as shown in **Table 2**.

Table 2. EXPre items scored Y and demonstrating high κ_{free}.

ALL	Students	CPD
Relaxation Tingling	Inner bodily flow Tingling Warmth or coolness	Being at ease Relaxation

Note: In addition, EXPre 'Inner bodily awareness' was scored Y with high κ_{free} in Pilot 3.

Those EXPost items with $\kappa_{free} > 0.7$ In the Pilots were 'Worry', 'Being blue' and 'Suppleness' (see **Table 10** below for how these were scored).

Reliability: Test-retest reliability

Short-term (1-7 week) test-retest reliability for EXPost in Pilots 1-4

Nine items demonstrated significant test-retest reliability (ρ or $\tau \geq 0.7$) between paired visits (visit 1 vs visit 2, etc.), in two or more Pilots.

Table 3. Items demonstrating significant short-term reliability (ρ or $\tau \geq 0.7$) in two or more Pilots, showing how many occurrences of significance were found, together with their Y and N scores and statistical significance of the Binomial test for proportion of these scores.

EXPost Item	Occurrences	*Total Y scores*	*Total N scores*	*p*
Inner bodily flow [0,1]	9	32	44	ns
Relaxation (4,4)	9	64	14	<0.001
Calmness (4,4)	6	56	22	<0.001
Sleepiness (1,4)	6	45	33	ns
Being at ease (4,2)	3	54	24	.001
Contentment (0,0)	3	38	38	ns
Inner bodily awareness (1,1)	3	37	40	ns
Nervousness [4,1]	2	10	66	<0.001
Restlessness (0,0)	2	14	63	<0.001

Note: Numbers in parentheses indicate in how many visits (0 to 4) these items occurred in the 'top five' list of Y responses, followed by the number of visits in which they appeared among the 'top five' asterisked items. Numbers in square brackets indicate in how many visits these items occurred in the 'top five' list of N responses, followed by the number of visits in which they appeared among the 'top five' DK responses. (No items displayed a mix of 'top five' Y and N scores).

Long-term (1 year) test-retest Reliability for those who participated in both Pilots 1 and 3

Complete agreement for all four participants on Y, N or DK scores did not occur for any EXPre or EXPost item.

No EXPre items demonstrated significant test-retest reliability. However, 'At ease', 'Calmness', 'Inner bodily awareness' and 'Relaxation' were each expected by the same three out of four participants in both Pilots.

EXPost 'Calmness' showed significant reliability in six out of eight possible comparisons (Pilot 1 visits 1-2 vs Pilot 3 visits 1-4), and 'Sleepiness' in three comparisons (ρ and $\tau = 1$, $p < 0.01$). In those comparisons for which reliability was significant, a change in Calmness was experienced by three out of four participants in all six comparisons, but a change in Sleepiness by only one out of four. Twelve other items showed the same reliability (ρ and $\tau = 1$, $p < 0.01$) in only one or two out of all eight comparisons, and 18 none. Thus only 21 out of a possible 256 comparisons showed significance.

Reliability: Split-half reliability

Table 4. Split-half reliability for EXPre and EXPost in the different cohorts.

EXP version	Correlation between halves	Cronbach's *alpha*	Mean inter-item correlation	Guttman split-half coefficient	*N* valid cases
EXPre (ALL)	0.818	[1] 0.865	0.289	0.898	208
		[2] 0.881	0.319		(2 excluded)
EXPre (All pilots)	0.882	[1] 0.879	0.295	0.934	25
		[2] 0.867	0.308		(0 excluded)
EXPre (Students)	0.819	[1] 0.873	0.294	0.898	129
		[2] 0.896	0.339		(0 excluded)
EXPre (CPD)	0.656	[1] 0.868	0.280	0.788	54
		[2] 0.770	0.167		(0 excluded)
EXPost (ALL v1)	0.956	[1] 0.975	0.708	0.977	208
		[2] 0.970	0.672		(2 excluded)
EXPost (All pilots v1)	0.909	[1] 0.907	0.391	0.952	25
		[2] 0.905	0.384		(0 excluded)
EXPost (Students)	0.967	[1]0.979	0.748	0.983	129
		[2] 0.974	0.709		(0 excluded)
EXPost (CPD)	0.926	[1] 0.971	0.675	0.961	54
		[2] 0.968	0.655		(0 excluded)

For EXPre, there was a small (<0.05) increase in mean inter-item correlation (IIC) between the two halves of the questionnaire for all cohorts except CPD, and for EXPost a small decrease in both Cronbach's *alpha* (≤0.005) and mean ICC (<0.04), for all cohorts. IIC variance in both EXPre and EXPost was small (≤0.03), but marginally greater for the second half in each cohort, the sole exception being for EXPost in the Pilot cohort.

However, these IIC and variance patterns were no longer evident when the order of items was randomised. Furthermore, differences in *alpha* and IIC between the two halves were numerically less in all cohorts when EXPre was randomised than when nonrandomised, and similarly for EXPost for most cohorts (except for CPD, with *alpha* being numerically greater as well for EXPost in the Pilots cohort).

These differences do suggest a small order effect, with some possible boredom ('questionnaire fatigue') in EXPre, but not EXPost. However, this interpretation is questionable and, given the size of the differences, they are probably negligible.

Split-half reliability (like inter-rater reliability) was good (>0.8) for the whole sample and for all except the CPD cohort (EXPre). (The difference in EXPre IIC between the two halves was most marked for this cohort)

Split-half reliability was consistently greater for EXPost than EXPre, with mean IIC sometimes considerably better for EXPost than EXPre for the corresponding cohort. In keeping with this finding, variance (for both halves taken together) was less for EXPost than EXPre in the whole sample and in all cohorts.

Results when the order of items was randomised were very similar. For EXPre, they differed by <10% from the results for nonrandomized item order for all values, except for the CPD cohort, where results differed by up to 33%; for EXPost, they differed by <7% for all values.

In this respect, stability was again more (variation less) for EXPost than EXPre. (Variance remains unchanged when item order is randomized.)

Reliability: Cronbach's *alpha*

Although the low inter-item correlations in the Table of split-half reliability results are mostly low (<0.7), the relatively high values for *alpha* shown there suggest it would be a useful exercise to calculate *alpha* for each complete rather than split questionnaire (*alpha* being mathematically equivalent to the average of all possible split-half estimates; Trochim, 2006).

As for split-half reliability, *alpha* for the whole questionnaire was higher for EXPost than EXPre in the corresponding cohorts.

Table 5. Cronbach's *alpha* for EXPre and EXPost in the different cohorts.

EXP version	Cronbach's *alpha*	Mean inter-item correlation (range)	*N* valid cases
EXPre (ALL)	0.929	0.293 (0.063 – 0.740)	208 (4 excluded)
EXPre (All pilots)	0.933	0.302 (-0.190 – 0.880)	25 (0 excluded)
EXPre (Students)	0.932	0.304 (0.040 – 0.719)	129 (0 excluded)
EXPre (CPD)	0.904	0.225 (-0.259 – 0.702)	54 (0 excluded)
EXPost (ALL v1)	0.986	0.684 (0.539 – 0.891)	208 (4 excluded)
EXPost (All pilots v1)	0.951	0.389 (-0.068 – 0.858	25 (0 excluded)
EXPost (Students)	0.988	0.724 (0.540 - 0.918)	129 (0 excluded)
EXPost (CPD)	0.983	0.650 (0.437 - 0.895)	54 (0 excluded)

Splitting the two questionnaires into 4 groups of 8 items alphabetically, *alpha* remained >0.7 for all groups except for EXPre items 17-24 in the CPD cohort, with IIC > 0.5 in 12 out of a possible 32 groups. Splitting the questionnaires into groups of 4 items, *alpha* remained >0.7 for only 33 of a possible 64 groups (IIC > 0.5 in 25 out of 64).

Cluster analysis

Formal cluster analysis

Clusters were derived on the basis of agreement between the different cohorts and the complete sample (see **Appendix III**). Removal of one or more equivocal items from the clusters did not greatly change their values of *alpha* or mean IIC, although *alpha* decreased as more items were removed, as expected. So, for example, for an average value of 0.963 for four variations of EXPost cluster 1, SD was only 0.0035.

Clusters extracted on the basis of positive IICs ≥0.5 and counts of negative IICs

For EXPre (Pilots cohort), out of a possible 528 correlations between items there were 66 (12.5%) with IIC ≥0.5 (including 14 ≥0.6, 7 ≥0.7 and 3≥0.8), all being positive. There were only 28 negative correlations in all (5.3%), with a mean negative IIC of -0.078 (SD 0.058), considerably less than that of the positive IICs (mean of all IICs was 0.296, SD 0.047). Clustering items with IICs ≥0.5 and ensuring those with a negative IIC were separated, resulted in four clusters (see **Appendix III** Group D, below).

Table 6. *Alpha* and mean IIC for fourfold clusters resulting from exploratory and confirmatory analysis

Basis of clusters	EXPre		EXPost	
	alpha	IIC	*alpha*	IIC
(1) Formal cluster analysis (from ALL and cohort data)	0.807	0.385	0.876	0.641
(2) Informally, from IICs ≥0.5 and negative correlations	0.731	0.313	0.922	0.685
(3a) 'Negative'	0.822	0.341	0.964	**0.747**
(3b) 'Positive'	**0.871**	0.339	**0.967**	0.696
(3c) Neither 'negative' nor 'positive'	0.802	0.289	0.954	0.679
(3d) 'Relaxation'	0.809	0.349	0.951	0.710
(3) Mean of the above	0.826	0.330	0.959	0.708
(4a) 'Bodily' feelings	**0.847**	0.299	**0.967**	0.694
(4b) 'Emotional' feelings	0.732	0.286	0.942	0.704
(4c) 'Mental' feelings	0.789	0.348	0.943	**0.707**
(4d) 'General' feelings	0.837	0.316	0.957	0.675
(4) Mean of the above	0.801	0.312	0.952	0.695

The main finding was that, regardless of how the clusters were developed, either from the data or from preconceived notions of how to group it, *alpha* and mean IIC for EXPost was consistently greater than for the same EXPre clusters.

None of the EXPre clusters resulted in acceptable values of IIC, although some *alpha* were >0.8. Lowest *alpha* was for the informally derived and 'emotional' clusters, lowest IIC for (3c) and, again, the 'emotional' clusters.

In general, for EXPost, confirmatory analysis gave better results (in terms of *alpha* and mean IIC) than exploratory analysis, although IIC for the informally derived clusters was greater than for (3c) or (4d).

Association between expectation and experience

Table 7A. Changes between EXPre and EXPost scores in the different cohorts.

Cohort	N→DK	Y→DK	N→Y	DK→DK	N→N	Y→Y	DK→Y	Y→N	DK→N
	–14	–7	–6	–5	–1	1	3	6	8
ALL	90	95	476	156	1691	1473	730	898	1480
Pilots (*N*=21)	9	13	145	27	512	459	374	238	778
CPD (*N*=54)	7	22	96	15	432	299	96	249	212
Students (*N*=129)	74	60	235	114	747	715	260	411	490
College 1 (*4 gps*)	44	30	101	87	452	268	140	205	316
College 2 (*3 gps*)	20	27	107	26	259	353	101	188	163

Table 7B. Summary of EXPre and EXPost Y and N score counts and ratios.

Cohort	Y→(Any)	N→(Any)	Y/N ratio	(Any)→Y	(Any)→N	Y/N ratio	Y Post/Pre	N Post/Pre
ALL	2466	2257	1.09	2679	4069	0.66	1.09	1.80
Pilots (*N*=21)	710	666	1.07	978	1528	0.64	1.38	2.29
CPD (*N*=54)	570	535	1.07	491	893	0.55	0.86	1.67
Students (*N*=129)	1186	1056	1.12	1210	1648	0.73	1.02	1.56

Figure 2. Graphical illustration of changes between EXPre and EXPost scores in different cohorts (left), and for students from different acupuncture training colleges (right).

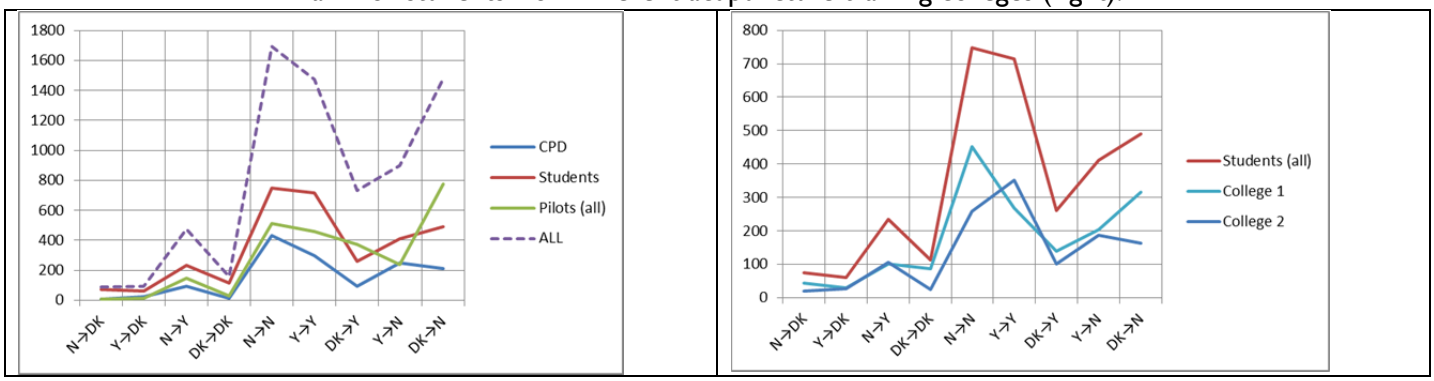


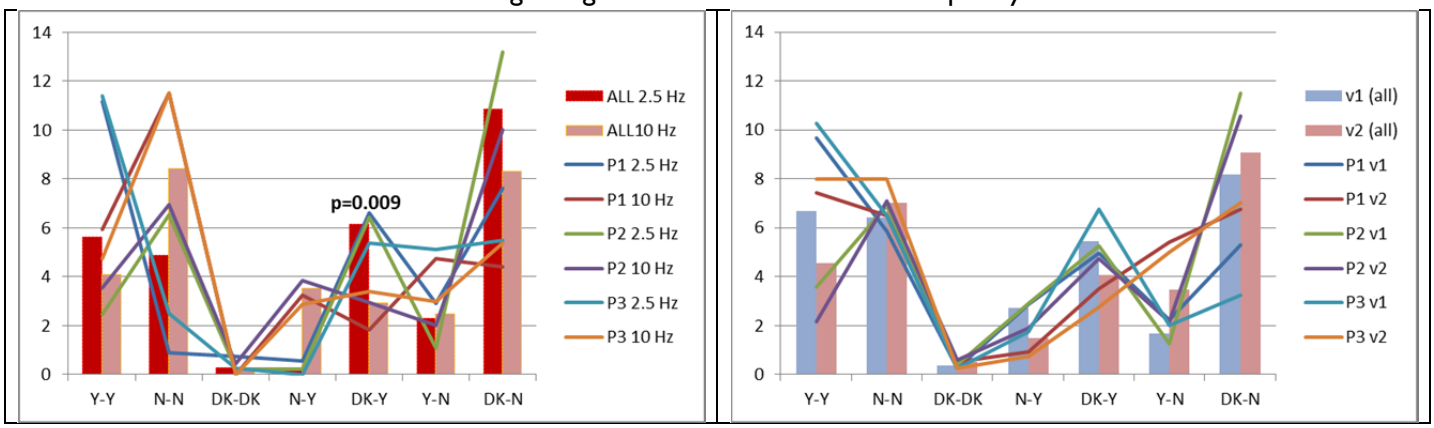
Figure 2 (left) shows that for the student and CPD cohorts, positive (Y→Y) and negative (N→N) expectations were fulfilled significantly more often than the other combinations occurred, in other words many of these participants **experienced what they expected** (Binomial, $p<0.001$ and $p=0.036$, respectively). This was also the case in the Pilot study, although here DK→N occurred significantly more frequently than either Y→Y or N→N ($p<0.001$) (in all cohorts, DK expectations became N experiences significantly more frequently than Y experiences; $p<0.001$).

However, comparing the total numbers of times expectations were met with the number of times they were not, for the complete sample and the Pilots, expectations were not fulfilled significantly more than they are (Binomial, $p<0.001$). However, for both the CPD and student cohorts there was a nonsignificant preponderance of fulfilled expectations.

Figure 2 (right) shows how results differed for students at two different acupuncture training colleges, with positive expectations fulfilled more at one ($p=0.001$), where there was also a nonsignificant preponderance of fulfilled expectations, and negative expectations more at the other ($p<0.001$), where there was also a nonsignificant preponderance of nonfulfilled expectations. It would be interesting to explore whether this reflects a difference in teaching methods.

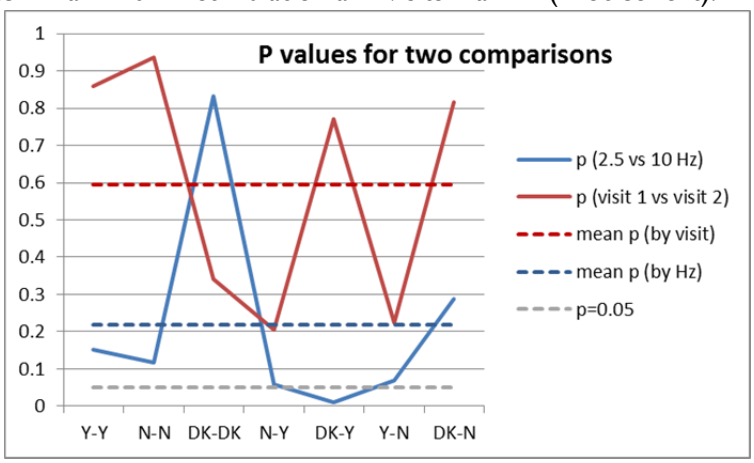
In Pilot studies 1 to 3, it was possible to compare results for different stimulation frequencies and over different visits. This is illustrated in **Figure 3**.

Figure 3. Left: Mean numbers of EXPre-EXPost score differences, comparing 2.5 Hz and 10 Hz sessions (across all visits). Right: Mean numbers of EXPre-EXPost score differences, comparing sessions (visit 1 vs visit 2), but disregarding the effects of stimulation frequency.



The Mann-Whitney U test showed only one significant difference (in DK→Y counts) between 2.5 Hz and 10 Hz stimulation, and no significant differences between visits. Visually, it appears that the differential effects of stimulation frequency are greater than those of visit. Examination of the actual p values for the Mann-Whitney U test for independent samples supports this impression (**Figure 4**).

Figure 4. Mann-Whitney U test p values for EXPre-EXPost score differences between 2.5 Hz and 10 Hz stimulation and visits 1 and 2 (Pilot cohort).



Most/least frequently found items

In order to simplify this presentation, only the most frequently found items are shown below.

Table 8. 'Top five' most frequently found EXPre items.

EXPre	Agreement over different cohorts or for complete sample
Yes (1)	Aliveness, Calmness, Inner bodily awareness, Inner bodily flow, Relaxation[a], Tension, Tingling[a]
No (2)	Being blue, Hunger[a], Nervousness, Worry
DK (3)	Contentment, Intestinal rumblings, Receptivity

Table 9. 'Top five' most frequently found EXPost items.

EXPost	Agreement over different cohorts or for complete sample
Yes (1 & *)	Aliveness, Being at ease, Calmness, Mental focus, Relaxation, Tingling
*	Aliveness, Being at ease, Calmness, Heaviness, Pain, Relaxation, Tension, Tingling, Warmth or coolness
No (2)	Being blue[a], Hunger[a], Intestinal rumblings[a], Restlessness[a], Worry[a]
DK (3)	Connectedness to others, Contentment, Physical vitality, Receptivity, Suppleness etc.

a. Items for which high κ_{free} was also found when scored in the same way.

Not all EXPre items showed significant Chi-square values in all cohorts (distribution of Y, N and DK significantly different from 1/3 each), but all EXPost items did (p between <0.001 and 0.015). This suggests that – for whatever reason – responses on feelings experienced differed from chance more than responses on feelings expected.

Table 10. Those items occurring most frequently in the EXPre→EXPost combinations of Y, N and DK listed in Table 7.

EXPre→EXPost	Agreement over different cohorts or for complete sample
N→DK [-14]	Being blue, Being in control, Receptive, Suppleness, Worry
Y→DK [-7]	Clarity, Contentment, Physical vitality. Suppleness, Tension
N→Y [-6]	At ease, Calmness, Contentment
DK→DK [-5]	Connectedness to others, Receptivity, Suppleness
N→N [-1]	Being blue, Hunger, Restlessness, Worry
Y→Y [1]	Aliveness, Calmness, Inner bodily flow, Tingling

DK→Y [3]	Being at ease, Peacefulness
Y→N [6]	Pain, Tension [etc.]
DK→N [8]	Excitement, Intestinal rumblings, Worry

Which feelings are expected or experienced could depend among other things on prior experience of acupuncture, or on tuition received during acupuncture training, for example. This could well be the case for items relating to 'Relaxation'. However, it is difficult on this basis to account for the prominence here of a term such as 'Aliveness', as although theoretically a nonspecific feeling resulting from a *qi*-based intervention (Ots, 1994), it is very unlikely that this term was used during tuition, and is not found in the literature on acupuncture responses.

Association with other trait and state measures

Few meaningful **trait** correlations were found (31 were significant, out of a possible 416, or 7.5%). For example, BFI 'Agreeableness' and 'Openness to experience' may contribute to the placebo effect (Kelley et al., 2009). However, although twelve significant correlations were found between the latter and various Y, N and DK total counts, there was only one significant correlation for 'Agreeableness' (with the number of DK responses to the 'Relaxation' EXPre cluster). Numbers of significant correlations for BAS-D (Drive), BAS-F (Fun seeking) and BAS-R (Reward responsiveness) were also in double figures, but difficult to interpret. Counter-intuitively, the single significant correlation for LOT-R was with number of Y→N EXPre-to-EXPost changes. Perhaps more predictably, MISS 'Stubborn opinionatedness' correlated positively with number of Y→Y changes (as did BAS/BAS 'Drive' and BFI 'Openness', among others).

Intriguingly, EXPre Y responses showed mostly positive significant correlations with the trait questionnaire scores, and all the N responses negative significant correlations (Binomial test significance, $p = 0.001$ and $p < 0.001$, respectively). The only exception to this was a negative correlation of BFI-E, 'Extraversion', with Y counts for the EXPre 'Relaxation' cluster (discussed below).

Assessing association between trait scales and individual EXPre and EXPost items did not yield useful results (of some 1600 possible associations, *eta* was > 0.7 for 951, too many to make any sensible selection). Those EXPre items items that appeared most frequently with *eta* > 0.7 were 'Clarity', 'Hunger', 'Peacefulness' and 'Receptivity'. The corresponding EXPost items were 'Tension' and 'Worry'.

The **state** scales showed 65 significant correlations (out of a possible 580, or 11.2%) with EXPre Y, N and DK, and the various EXPre→EXPost changes, 45 being positive and 20 negative. The scale with most (11) positive correlations was VAS-R (Relaxation immediately after stimulation, VAS-R_{post}, or over the preceding month, VAS-R_{mth} – either before first visit or at one-month follow up), that with most (11) negative correlations POMS-T ('Tension-Anxiety'). **Table 11** shows some examples.

Table 11. Some potentially meaningful positive and negative correlations ≥ 0.4

Scale	EXPre or EXPre→EXPost	Spearman's *rho*	Significance
VAS-R_{mth} (before visit 1)	EXPre 'Negative' Y	0.554	p<0.001
VAS-R_{mth} (before visit 1)	Y→Y	0.497	p<0.001
VAS-R_{post} (after stimulation)	EXPre 'Relaxation' N	-0.467	p=0.001
VAS-R_{post} (after stimulation)	DK→Y	0.419	p=0.003
VAS-R_{mth} (before follow up)	EXPre 'Relaxation' N	-0.409	p=0.002
POMS-T (Tension …) after stimulation	EXPre 'Relaxation' N	0.468	p=0.001
POMS-T (Tension …) after stimulation	N→Y	0.416	p=0.003
POMS-T (Tension …) after stimulation	DK→N	-0.430	p=0.002
POMS-V (Vigour …) at follow up	N→N	-0.520	p=0.002
POMS-C (Confusion …) pre stimulation	DK→DK	0.458	p=0.001
POMS-C (Confusion …) at follow up	DK→DK	0.416	p=0.018

All 11 negative correlations for POMS-T were with EXpre or EXPre-to-EXPost changes that included DK counts, suggesting perhaps that those who felt less anxious also felt more able to indicate a DK rather than a forced Y or N

score (whether for all items taken together, or for the 'Negative', 'Positive' or 'Relaxation' clusters). Because of the small numbers involved, however, any such interpretations drawn from these correlations can only be tentative.

Assessing associations between state scales and individual EXPre and EXPost items yielded 113 with *eta* > 0.7 out of 1792 (6.3%), 58 for EXPre (26 items) and 55 for EXPost (21 items), as shown in **Table 12**.

Table 12. State scales with mean *eta* correlation ratio values > 0.7 for association with EXPre and EXPost items (VAS-R_{pre} was used to assess current state of relation immediately prior to stimulation).

	EXPre	mean *eta*	EXPost	mean *eta*
POMS-V_{pre}	Aliveness Clarity Heaviness Intestinal rumblings	0.721	Clarity Mental energy Mental focus	0.750
POMS-TMD_{pre}	Intestinal rumblings Peacefulness	0.719	Excitement	0.749
POMS-TMD_{post}	(na)	–	Connectedness to others Excitement Tension	0.731
PSS-10	Aliveness	0.721	Excitement	0.714
VAS-R_{mth}	all except: Being blue Cheerfulness Connectedness to others Contentment Inner bodily awareness Pain Relaxation	0.803	all except: Being blue Being spaced out Cheerfulness Intestinal rumblings Physical vitality Relaxation Sensory acuteness Sleepiness Suppleness Tingling Warmth or coolness Worry	0.763
VAS-R_{pre}	all except: Cheerfulness Connectedness to others Contentment Inner bodily awareness Pain Relaxation	0.808	all except: Being blue Being spaced out Cheerfulness Intestinal rumblings Relaxation Sensory acuteness Sleepiness Suppleness Tingling Warmth or coolness Worry	0.756
VAS-R_{post}	(na)	–	Clarity Connectedness to others Excitement Heaviness Hunger Nervousness	0.736

Without knowing whether *eta* > 0.7 indicates a positive or negative association, the above POMS associations are not easy to interpret, although it is tempting to suggest that greater POMS 'Vigour-Activity' prior to stimulation results in the experience of changes in clarity, mental energy and focus, for example.

Of more interest is the *lack* of association between the VAS-R_{mth}, VAS-R_{pre} and VAS-R_{post} scales, intended to measure 'Relaxation' (experienced over the past month, or currently – before or after stimulation), and the expectation or experience of change in relaxation during stimulation. This suggests either that VAS-R does not in fact measure relaxation, or that the *degree* of relaxation experienced is not necessarily related to a *change* in the feeling of relaxation experienced.

Greater *eta* for EXPre than EXPost VAS-R suggests that pre-existing relaxation has more effect on expectation of changes in feelings than on their subsequent experience.

Influence of treatment aspects on reported feelings (Pilots cohort)

1. Participant ID

For EXPre, *Lambda* =1 ($p<0.001$) for all items, indicating a great variety of responses among individuals.

Table 13. EXPost results for *lambda* and its significance(values and significance for All Pilots; significance for Pilots 1 to 3 individually).

EXPost item	*Lambda*	significance				EXPost item	*Lambda*	significance			
		All	P1	P2	P3			All	P1	P2	P3
Aliveness	0.400	**	*	**	ns	Mental energy	0.629	**	ns	**	ns
Being at ease	0.593	**	**	**	ns	Mental focus	0.649	**	ns	*	ns
Being blue	0.429	ns	**	ns	ns	Nervousness	0.231	ns	ns	ns	ns
Being in control	0.500	*	ns	**	ns	Pain	0.333	ns	**	ns	ns
Spaced out	0.579	**	ns	**	ns	Peacefulness	0.500	**	*	*	*
Calmness	0.680	**	**	**	*	Physical vitality	0.436	**	*	ns	ns
Cheerfulness	0.483	**	**	ns	ns	Receptivity	0.548	**	**	*	ns
Clarity	0.515	**	*	**	ns	Relaxation	0.706	**	ns	*	*
Connectedness	0.522	**	*	ns	ns	Restlessness	0.250	ns	ns	ns	ns
Contentment	0.675	**	**	**	ns	Sensory acuteness	0.647	**	ns	ns	*
Excitement	0.267	ns	ns	ns	ns	Sleepiness	0.647	**	ns	**	*
Heaviness	0.641	**	*	*	ns	Suppleness	0.364	*	**	ns	ns
Hunger	0.188	ns	ns	ns	ns	Tension	0.333	ns	ns	ns	ns
Inner awareness	0.775	**	**	**	*	Tingling	0.559	**	*	ns	ns
Inner flow	0.784	**	ns	*	*	Warmth-coolness	0.517	**	**	ns	ns
Intestinal rumblings	0.433	*	*	ns	ns	Worry	0.250	ns	**	ns	ns

** $p<0.01$; * $p<0.05$; ns not significant.

In All pilots, least variation between participants was thus found for 'Being blue', 'Excitement', 'Hunger', 'Nervousness', 'Pain', 'Restlessness', 'Tension' and 'Worry', along with 'Being in control', 'Intestinal rumblings' and 'Suppleness'. It is instructive to compare those Items showing significance across three or four columns with the items showing short-term test-retest reliability and most frequently found items (above, **Tables 3** and **9**).

Whether an item was asterisked or not also showed variation across participants for All pilots (*lambda* = 0.615 $p=0.002$), in Pilot 1 (*lambda* = 1, $p<0.01$), but not in Pilots 2 or 3 (*lambda* = 0.5, ns).

2. Visit

Lambda was significant ($p=0.04$) but very low (0.216) for only one EXPost item ('Mental focus').

3. Stimulation location

Lambda was significant ($p=0.017$) but very low (0.216) for only one EXPost item ('Inner bodily flow').

4. Stimulation frequency

Lambda was not significant for any EXPost items.

Thus, in contrast to the associations between expectation and experience, which appeared to be significantly dependent on Stimulation frequency, but not on Visit (above), it is unlikely that Visit order, Stimulation location or Stimulation frequency have any meaningful association with particular EXPost items. Furthermore, there appeared to be no significant dependence of the mean number of items asterisked on either visit or stimulation frequency, although more items were asterisked following a first stimulation session, but fewer after subsequent sessions.

Possible correlations with stimulation amplitude and duration were not analysed.

Summary of salient results

Content validity was strictly found for only two items ('Pain' and 'Relaxation'). When criteria were relaxed, acceptable I-CVI and S-CVI_{AV-UA} were found for longer lists of items, particularly for women rather than men, and more so when non-practitioner researchers were excluded from analysis. Results, together with an ensuing discussion of the term 'nonspecific', will be reported elsewhere.

Inter-rater reliability (κ_{free}) was low, as expected, achieving significance only for a few EXPost items in the Pilot cohort. In general, it was higher for EXPost than EXPre. Across cohorts, only two EXPre items were expected with high κ_{free} ('Relaxation' and 'Tingling'). Students demonstrated less EXPost agreement than experienced practitioners.

Test-retest reliability (short-term, across visits) was significant for nine EXPost items. Of these, three were scored 'Yes' significantly more often than 'No' ('Being at ease', 'Calmness' and 'Relaxation'), two scored 'No' significantly more often than 'Yes' ('Nervousness' and 'Restlessness').

Test-retest reliability (long-term, across Pilots) was not significant for any EXPre items, although 'At ease', 'Calmness', 'Inner bodily awareness' and 'Relaxation' were each expected by the same three out of four participants in both Pilots. Significant reliability was found most often for EXPost 'Calmness'.

Split-half reliability (EXPre) was > 0.8 for all except the CPD cohort, and consistently greater for EXPost than EXPre. There was a small increase in mean inter-item correlation (IIC) between the two halves of the EXPre questionnaire for all cohorts except CPD, and for EXPost a small decrease in both Cronbach's *alpha* and mean ICC for all cohorts. These differences, although small, suggest possible 'questionnaire fatigue' in response to EXPre, but not EXPost.

Cronbach's *alpha* was > 0.9 for all cohorts, and consistently higher for EXPost than EXPre in corresponding cohorts. Dividing the questionnaires into four equal subquestionnaires, *alpha* remained > 0.7 for all but one subquestionnaire in the CPD cohort.

Cluster analysis. Regardless of how the clusters were developed, either from the data or from preconceived notions of how to group it, *alpha* and mean IIC for EXPost was consistently greater than for the same EXPre clusters.

None of the EXPre clusters resulted in acceptable values of IIC, although some *alpha* were >0.8. In general, for EXPost, confirmatory analysis gave better results (in terms of *alpha* and mean IIC) than exploratory analysis

Expectation and experience. For the student and CPD cohorts, positive (Y→Y) and negative (N→N) expectations were fulfilled significantly more often than the other combinations occurred, in other words many of these participants **experienced what they expected**. This was also the case in the Pilot study, although here DK→N occurred significantly more frequently than either Y→Y or N→N.

In all cohorts, EXPre Y and N scores occurred with similar frequency (mean Y/N ratio 1.09, range 1.07 to 1.12). In contrast, EXPost N scores occurred significantly more frequently than Y scores (mean Y/N ratio 0.65, range 0.55 to 0.73). Mean EXPost/EXPre Y ratio was 1.09 (range 0.86-1.38), whereas EXPost/EXPre N ratio was 1.83 (range 1.56-2.29).

A variation in response between students at two acupuncture training colleges suggests a possible difference in teaching methods. N→Y and DK→Y responses may also vary with stimulation frequency.

Table 14. *Most frequently found items* scored 'Yes', asterisked, or scored Y in both EXPre and EXPost.

Item	EXPre (Y)	EXPost (Y)	EXPost (*)	Y→Y
Aliveness	Y	Y	*	Y→Y
Being at ease[b]	–[d,e]	Y	*	–
Calmness[b]	Y[d,e]	Y[e]	*	Y→Y
Heaviness[b]	–	–	*	–
Inner bodily awareness	Y[c]	–	–	–
Inner bodily flow	Y	–	–	Y→Y
Mental focus[b]	–	Y	–	–
Pain[a,b]	–	–[f]	*	–
Relaxation[a,b]	Y[c,d,e]	Y	*	–

Tension	Y	Y[f]	*	–
Tingling	Y[e]	Y	*	Y→Y
Warmth or coolness[b]	–	–	*	–

Overlaps: a. Lawshe's Content Validity Ratio (CVR) and Lynn's Content Validity Index (CVI) acceptable; b. CVI 'excellent' when 'useful' items scored as Essential; c. Inter-rater reliability; d. Short-term test-retest reliability; e. Long-term test-retest reliability; f. Least variation between participants (lambda non-significant).

Association with other trait and state measures

Few meaningful trait correlations were found, although EXPre Y responses showed mostly positive significant correlations with the trait questionnaire scores, and all the N responses negative significant correlations. The only exception to this was a negative correlation of BFI-E, 'Extraversion', with Y counts for the EXPre 'Relaxation' cluster.

The state scale with most positive correlations with EXPre Y, N and DK, and the various EXPre→EXPost changes, was VAS-R (Relaxation after stimulation, or over the preceding month – either before first visit or at one-month follow up), that with most (11) negative correlations POMS-T ('Tension-Anxiety'). All 11 negative correlations for POMS-T were with EXpre or EXPre-to-EXPost changes that included DK counts.

A lack of association between VAS-R and the EXPre and EXPost 'Relaxation' item suggests that the *degree* of relaxation experienced is not necessarily related to a *change* in the feeling of relaxation experienced, although pre-existing relaxation may have more effect on expectation of changes in feelings than on their subsequent experience.

Influence of treatment aspects on reported feelings. Whereas both EXPre and EXPost responses varied considerably among participants, suggesting the possibility of different response styles, it appears unlikely that Visit order, Stimulation location or Stimulation frequency have any meaningful association with particular EXPost items.

Discussion

As with many interventions (Bingel et al., 2011; Finch et al., 2005; Flood et al., 1993; Leedham et al., 1995), positive expectation of acupuncture effects may correlate significantly with perceived or actual clinical outcome (Bausell et al., 2005; Kalauokalani et al., 2001; Linde et al., 2007; Vase et al., 2013), enhancing acupuncture analgesia, for example (Kong et al., 2009b), although if expectation of benefit is too high, outcome may be less favourable (So, 2002), and negative expectation may adversely affect response (Chae et al., 2008). One research group has found not only that positive expectation may significantly predict the response to *verum* but not placebo acupuncture (Wasan et al., 2010), but that even when high expectation enhances subjective reports of analgesia equally for both *verum* and sham acupuncture, objective (fMRI) measures indicate differences in underlying mechanisms between the two interventions (Kong et al., 2009a; cf. Pariente et al., 2005).

In keeping with these findings, our results do suggest that expectations of change are often followed by the experience of such change, and that expectation of no change may similarly be followed by no change being experienced.

Interestingly, dental patients have been shown to accurately anticipate the *pattern* of sensations involved in treatment (even if not experienced before), but tend to expect more intense sensations and greater discomfort and apprehension than they actually experience (Lindsay et al. 1984).

In a way the responses of our participants parallel this finding, with mean EXPre Y/N and EXPost/ EXPre Y ratios of 1.09, but an EXPost/EXPre N ratio of 1.83. Thus, although the pattern of Y responses was similar before and after stimulation, there were many more N responses after.

This is in line with the repeated finding that reliability (consistency of response) of the various types tested is greater in EXPost than EXPre.

The negative correlation of BFI-E, 'Extraversion', with Y counts for the EXPre 'Relaxation' cluster could perhaps be accounted for on the basis of the hypothesis that extraverts may seek external stimulation due to chronic under-arousal and may therefore find it more difficult to relax than already over-aroused introverts [Wilderdom, 2003; cf.

Leboeuf, 1977). However, the result, as one among many, should be considered with caution, and only as a basis for further investigation with more cases.

Limitations and further investigations

There are inherent difficulties in attempting to assess the nonspecific expectations and experiences of treatment, not least because these are themselves malleable, liable to change in response to many factors, and in ways of which we may not be fully aware (Stone et al., 2005). It would therefore be prudent to explore the use of the EXPre and EXPost questionnaires in other settings, with participants drawn from different populations (for example, those who have not had prior experience of acupuncture-related interventions), and for other treatments (including, perhaps, some not related to acupuncture or other forms of 'energy medicine'). As far as possible, more careful attention should be paid to accounting for or eliminating potential confounders (such as ambiguities in questionnaire wording, participant interaction and distraction in group settings, or duration of stimulation). It may indeed be fruitful to compare expectations and experiences of different treatments, such as manual acupuncture, EA and TEAS, as well as 'sham' versions of these.

In terms of analysis, multivariate methods could be employed to distinguish between the effects of different factors, and tests for multiple comparisons (e.g. the Bonferroni correction) should be conducted. To confirm results found for small cohorts, where appropriate a resampling method (such as Bootstrap) should be used.

Acknowledgments

To Robert Kozarski and Tim Watson of the University of Hertfordshire, for trying to keep us on the straight and narrow; to those who gave freely of their advice early in the development of the questionnaires and to those who took part more recently in the content validity survey, in particular to Stephen Birch for his wise advice on using the term 'nonspecific'; to Klaus Linde and Louise Percival for drawing attention to ambiguities in question wording; to Richard Harrison and Kevin August for keeping DM's computer limping through a bad case of hard drive failure while preparing this presentation (and to DM's wife for her postoperative care at the same time); and of course to the acupuncturists and other health practitioners who gave up considerable amounts of their time to take part in our EEG/HRV study, and those acupuncture students and practitioners who completed questionnaires as part of their learning about electroacupuncture.

References

Bausell RB, Lao L, Bergman S, Lee WL, Berman BM. Is acupuncture analgesia an expectancy effect? Preliminary evidence based on participants' perceived assignments in two placebo-controlled trials. Evaluation and the Health Professions. 2005;28(1):9-26

Bingel U, Wanigasekera V, Wiech K, Ni Mhuircheartaigh R, Lee MC, Ploner M, Tracey I. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanil. Science Translational Medicine. 2011;3(70):70ra14

Blasche G, Marktl W, Eisenwort B, Skolka A, Pichlhöfer O. The Treatment Experience Questionnaire: Development and Validation of a Questionnaire Assessing the Individual's Emotional, Perceptual, and Cognitive Reactions to Alternative, Physical, and Dental Treatments. Forschende Komplementärmedizin. 2013;20(3):205-12

Borgatti S. 2008. Cronbach's alpha. Gatton College of Business & Economics, University of Kentucky. http://www.analytictech.com/ba762/handouts/alpha.htm [accessed 22 July 2013]

Brennan RL, Prediger DJ. Coefficient Kappa: Some uses, misuses, and alternatives. Educational and Psychological Measurement. 1981;41(3):687-99

Bryman A, Cramer D. 2001. Quantitatve Data Analysis with SPSS Release 10 for Windows: A guide for social scientists. Hove, UK: Routledge

Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. Journal of Personality and Social Psychology. 1994;67(2):319-33

Chae Y, Kim SY, Park HS, Lee H, Park HJ. Experimentally manipulating perceptions regarding acupuncture elicits different responses to the identical acupuncture stimulation. Physiology and Behavior. 2008;95(3):515-20

Cohen S, Williamson G. 1988. Perceived stress in a probability sample of the United States. In: Spacapam S, Oskamp S. (eds.). The Social Psychology of Health: Claremont Symposium on applied psychology. Newbury Park, CA: Sage, pp. 31-67

DeVellis RF. 1991. Scale Development: Theory and applications. Newbury Park, CA: Sage

Everitt B. 1980. Cluster Analysis. London: Heinemann Educational

Finch EA, Linde JA, Jeffery RW, Rothman AJ, King CM, Levy RL. The effects of outcome expectations and satisfaction on weight loss and maintenance: correlational and experimental analyses--a randomized trial. Health Psychology. 2005;24(6):608-16

Flood AB, Lorence DP, Ding J, McPherson K, Black NA. The role of expectations in patients' reports of post-operative outcomes and improvement following therapy. Medical Care. 1993;31(11):1043-56

Geers AL, Wellman JA, Fowler SL, Rasinski HM, Helfer SG. Placebo expectations and the detection of somatic information. Journal of Behavioral Medicine. 2011;34(3):208-17

George D, Mallery P. 2003. SPSS for Windows Step by Step: A simple guide and reference 11.0 update. (4th ed.). Boston, MA: Allyn and Bacon

John OP, Naumann LP, Soto CJ, Paradigm shift to the integrative Big Five trait taxonomy: history, measurement, and conceptual issues. In: John OP, Robins RW, Pervin LA. (eds.). Handbook of Personality. Theory and research. (3rd ed.). New York: Guilford Press, pp. 114-58

Johnson JE. Effects of accurate expectations about sensations on the sensory and distress components of pain. Journal of Personality and Social Psychology. 1973;27(2):261-75

Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. Spine. 2001;26(13):1418-24

Kelley JM, Lembo AJ, Ablon JS, Villanueva JJ, Conboy LA, Levy R, Marci CD, Kerr CE, Kirsch I, Jacobson EE, Riess H, Kaptchuk TJ. Patient and practitioner influences on the placebo effect in irritable bowel syndrome. Psychosomatic Medicine. 2009;71(7):789-97

Kerr CE, Shaw JR, Conboy LA, Kelley JM, Jacobson E, Kaptchuk TJ. Placebo acupuncture as a form of ritual touch healing: A neurophenomenological model. Consciousness and Cognition. 2011; 20(3): 784-91

Kim Y, Park J, Lee H, Bang H, Park HJ. Content validity of an acupuncture sensation questionnaire. Journal of Alternative and Complementary Medicine. 2008;14(8):957-63

Kong J, Fufa DT, Gerber AJ, Rosman IS, Vangel MG, Gracely RH, Gollub RL. Psychophysical outcomes from a randomized pilot study of manual, electro, and sham acupuncture treatment on experimentally induced thermal pain. Journal of Pain. 2005;6(1):55-64

Kong J, Gollub R, Huang T, Polich G, Napadow V, Hui K, Vangel M, Rosen B, Kaptchuk TJ. 2007 Acupuncture de qi, from qualitative history to quantitative measurement. Journal of Alternative and Complementary Medicine. 2007;13(10):1059-70

Kong J, Kaptchuk TJ, Polich G, Kirsch I, Vangel M, Zyloney C, Rosen B, Gollub R. Expectancy and treatment interactions: a dissociation between acupuncture analgesia and expectancy evoked placebo analgesia. Neuroimage. 2009;45(3):940-9 [Kong et al., 2009a]

Kong J, Kaptchuk TJ, Polich G, Kirsch I, Vangel M, Zyloney C, Rosen B, Gollub RL. An fMRI study on the interaction and dissociation between expectation of pain relief and acupuncture treatment. Neuroimage. 2009;47(3):1066-76 [Kong et al., 2009b]

Kotov RI, Belmman SB, Watson DB. 2004. Multidimensional Iowa Suggestibility Scale (MISS) Brief Manual. Stony Brook School of Medicine. http://medicine.stonybrookmedicine.edu/system/files/MISSBriefManual.pdf [accessed 28 May 2011]

Lawshe CH. A quantitative approach to content validity. Personnel Psychology. 1975;28(4):563-75

Leboeuf A. The effects of EMG feedback training on state anxiety in introverts and extraverts. Journal of Clinical Psychology. 1977;33(1):251-3

Leedham B, Meyerowitz BE, Muirhead J, Frist WH. Positive expectations predict health after heart transplantation. Health Psychology. 1995;14(1):74-9

Linde K, Witt CM, Streng A, Weidenhammer W, Wagenpfeil S, Brinkhaus B, Willich SN, Melchart D. The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain. Pain. 2007;128(3):264-71

Linde K, Niemann K, Schneider A, Meissner K. How large are the nonspecific effects of acupuncture? A meta-analysis of randomized controlled trials. BMC Med. 2010 Nov 23;8:75

Lindsay SJ, Wege P, Yates J. Expectations of sensations, discomfort and fear in dental treatment. Behaviour Research and Therapy. 1984;22(2):99-108

Lynn MR. Determination and quantification of content validity. Nursing Research. 1986;35(6):382-5

Mayor D. 2011. Elemental souls and vernacular qi: some attributes of what moves us. In: Mayor D, Micozzi MS. (eds.). Energy Medicine East and West: A natural history of qi. Edinburgh: Churchill Livingstone, pp. 24-47

Morris A. 2007. Exploration of the phenomena of qi. A qualitative study of patient/practitioner experience. Dissertation submitted in partial fulfilment of the requirements for the BA (Honours) in Traditional Acupuncture. Warwick, UK: College of Traditional Acupuncture, in collaboration with Oxford Brookes University

Nunnally JC, Bernstein IH. 1994. Psychometric Theory. New York: McGraw-Hill

Ots T. 1994. The silenced body – the expressive Leib: on the dialectic of mind and life in Chinese cathartic healing. In: Csordas TJ. (ed). Embodiment and Experience: The existential ground of culture and self. Cambridge University Press, Cambridge, pp. 116-36

Pach D, Hohmann C, Lüdtke R, Zimmermann-Viehoff F, Witt CM, Thiele C. German translation of the Southampton Needle Sensation Questionnaire: use in an experimental acupuncture study. Forschende Komplementärmedizin. 2011;18(6):321-6

Pacheco-López G, Engler H, Niemi MB, Schedlowski M. Expectations and associations that heal: Immunomodulatory placebo effects and its neurobiology. Brain, Behavior, and Immunity. 2006;20(5):430-46

Pariente J, White P, Frackowiak RS, Lewith G. Expectancy and belief modulate the neuronal substrates of pain treated by acupuncture. Neuroimage. 2005;25(4):1161-7

Park H, Park J, Lee H, Lee H. Does Deqi (needle sensation) exist? American Journal of Chinese Medicine. 2002;30(1):45-50

Park J, Park H, Lee H, Lim S, Ahn K, Lee H. Deqi sensation between the acupuncture-experienced and the naïve: a Korean study II. American Journal of Chinese Medicine. 2005;33(2):329-37

Peters D. (ed.). 2001. Understanding the Placebo Effect in Complementary Medicine: Theory, practice and research. London: Churchill Livingstone

Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing and Health. 2007;30(4):459-67

Randolph JJ. 2008. Online Kappa Calculator. http://justus.randolph.name/kappa [accessed 16 Dec 2011]

Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a re-evaluation of the Life Orientation Test. Journal of Personality and Social Psychology. 1994;67(6): 1063-78

Shacham S. A shortened version of the Profile of Mood States. Journal of Personality Assessment. 1983;47(3):305-6

So DW. Acupuncture outcomes, expectations, patient-provider relationship, and the placebo effect: implications for health promotion. American Journal of Public Health. 2002;92(10):1662-7

Stone DA, Kerr CE, Jacobson E, Conboy LA, Kaptchuk TJ. Patient expectations in placebo-controlled randomized clinical trials. Journal of Evaluation in Clinical Practice. 2005;11(1):77-84

Trochim WMK. 2006. Types of reliability. Web Center for Social Research Methods. http://www.socialresearchmethods.net/kb/reltypes.php [accessed 22 July 2013]

Vase L, Baram S, Takakura N, Yajima H, Takayama M, Kaptchuk TJ, Schou S, Jensen TS, Zachariae R, Svensson P. Specifying the nonspecific components of acupuncture analgesia. Pain. 2013 May 23. pii: S0304-3959(13)00231-5

Vearncombe H. 2007. External signs of internal movement of qi or energetic change in acupuncture. Research and reflective practice. BSc (Hons) Dissertation. Reading, UK: College of Integrated Chinese Medicine

Vincent CA, Richardson PH, Black JJ, Pither CE. The significance of needle placement site in acupuncture. Journal of Psychosomatic Research. 1989;33(4):489-96.

Walach H, Jonas WB. Placebo research: the evidence base for harnessing self-healing capacities. Journal of Alternative and Complementary Medicine. 2004;10 Suppl 1:S103-12

Warrens, MJ. 2010. Inequalities between multi-rater kappas. Advances in Data Analysis and Classification. 2010; 4(4):271-86

Wasan AD, Kong J, Pham LD, Kaptchuk TJ, Edwards R, Gollub RL. The impact of placebo, psychopathology, and expectations on the response to acupuncture needling in patients with chronic low back pain. Journal of Pain. 2010;11(6):555-63

White P, Bishop F, Hardy H, Abdollahian S, White A, Park J, Kaptchuk TJ, Lewith GT. Southampton needle sensation questionnaire: development and validation of a measure to gauge acupuncture needle sensation. Journal of Alternative and Complementary Medicine. 2008 May;14(4):373-9

Wilderdom. 2003. Biological processes in personality. Wilderdom. A project in natural living & transformation. http://www.wilderdom.com/personality/L7-4BiologicalProcessesPersonality.html [accessed 23 Jun 2013]

Wilson FR, Pan W, Schumsky DA. Recalculation of the critical values for Lawshe's content validity ratio. Measurement and Evaluation in Counseling and Development. 2012; 45(3):197-210

Yu DT, Jones AY, Pang MY. Development and validation of the Chinese version of the Massachusetts General Hospital Acupuncture Sensation Scale: an exploratory and methodological study. Acupuncture in Medicine. 2012;30(3):214-21

Yurdugül H. Minimum sample size for Cronbach's coefficient alpha: a Monte-Carlo study. Hacettepe University Journal of Education. 2008; 35:397-405

Appendix I. Items used in EXPre and EXPost questionnaires

Aliveness	Connectedness with others	Mental energy	Restlessness
Being at ease	Contentment	Mental focus	Sensory acuteness
Being blue	Excitement	Nervousness	Sleepiness
Being in control	Heaviness	Pain	Suppleness
Being spaced out	Hunger	Peacefulness	Tension
Calmness	Inner awareness	Physical vitality	Tingling
Cheerfulness	Inner flow	Receptivity	Warmth or coolness
Clarity	Intestinal rumblings	Relaxation	Worry

Appendix II. Additional items used in Content validity survey

Comfort	Harmony	Optimism	Satisfaction
Empowerment	Looseness	Positivity	Vibration
Expansion	Melting	Pulsation	[& Warmth or coolness
Floating	Numbness	Relief	split into 2 items]

Appendix III. Clusters

Hypothesised clusters for confirmatory analysis

Group A. Polarity style

[1] Items which might be construed as 'negative' in some way
[2] Items which might be construed as 'positive' in some way
[3] Items which might be construed as neither 'negative' nor 'positive'

[1] 'Negative' items (9)	[2] 'Positive' items (13)	[3] 'Neutral' items (10)
Being blue	Aliveness	Connectedness to others
Being spaced out	Being at ease	Inner bodily awareness
Heaviness	Being in control	Inner bodily flow
Hunger	Calmness	Intestinal rumblings
Nervousness	Cheerfulness	Receptivity
Pain	Clarity	Sensory acuteness
Restlessness	Contentment	Sleepiness
Tension	Excitement	Suppleness
Worry	Mental energy	Tingling
	Mental focus	Warmth or coolness
	Peacefulness	
	Physical vitality	
	Relaxation	

Group B. Feeling style

[4] 'bodily' feelings
[5] 'emotional' feelings
[6] 'mental' feelings
[7] 'general' feelings (interpretable as any of 'bodily', 'emotional' or 'mental'

[4] 'Bodily' feelings (13)	[5] 'Emotional' feelings (6)	[6] 'Mental' feelings (7)	[7] 'General' feelings (11)
Hunger	Being blue	Being spaced out	Aliveness
Inner bodily awareness	Cheerfulness	Clarity*	Being at ease
Inner bodily flow	Clarity*	Mental energy	Being in control
Intestinal rumblings	Nervousness*	Mental focus	Calmness
Nervousness*	Receptivity*	Receptivity*	Connectedness to others
Pain	Worry	Restlessness*	Contentment
Physical vitality		Sleepiness*	Excitement
Restlessness*			Heaviness

Sensory acuteness Sleepiness* Suppleness Tingling Warmth or coolness			Peacefulness Relaxation Tension

Asterisked items occur in more than one cluster.

[8] Items considered to relate to the construct 'relaxation'

[8] 'Relaxation' items (8)
Being at ease
Calmness
Contentment
Intestinal rumblings
Peacefulness
Relaxation
Sleepiness
Warmth or coolness

Clusters derived from data in exploratory analysis

Group C. EXPre and EXPost clusters derived using agglomerative hierarchical fourfold clustering method based on Ward's method, with a chi-squared measure for categorical (count) data, allowing a range of solutions (3-10).

EXPre

Cluster 1 (7)	Cluster 2 (5)	Cluster 3 (15-16)	Cluster 4 (5-6)
Aliveness Cheerfulness Clarity Excitement Mental energy Mental focus Physical vitality	Being at ease Calmness Contentment Peacefulness Relaxation	Being blue Being in control Being spaced out* Connectedness with others Heaviness Hunger Nervousness Pain Receptivity Restlessness Sleepiness Suppleness Tension Tingling Warmth or coolness Worry	Being in control Being spaced out* Inner bodily awareness Inner bodily flow Intestinal rumblings Sensory acuteness

Items asterisked indicate possible alternative cluster allocations.

EXPost

Cluster 1 (10-18)	Cluster 2 (5-8)	Cluster 3 (7-12)	Cluster 4 (2)
Aliveness Being in control* Being spaced out* Cheerfulness Clarity Connectedness with others* Contentment Excitement* Heaviness* Inner bodily awareness Inner bodily flow Intestinal rumblings*	Being at ease Being in control* Calmness Connectedness with others* Excitement* Peacefulness Relaxation Warmth or coolness	Blue Being spaced-out* Heaviness* Hunger Intestinal rumblings* Nervousness Pain Restlessness* Sleepiness* Suppleness Tension Worry	Receptivity Tingling

Mental energy Mental focus Physical vitality Restlessness* Sensory acuteness Sleepiness*			

Items asterisked indicate possible alternative cluster allocations.

Group D. EXPre clusters derived on the basis of positive IICs $\geq$ 0.5 and numbers of negative IICs

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Being at ease Being blue Calmness Contentment Inner bodily awareness Inner bodily flow Peacefulness Relaxation Tension Warmth or coolness	Aliveness Being spaced out Clarity Excitement Intestinal rumblings Mental energy Mental focus Nervousness Pain Physical vitality Restlessness Sleepiness Worry	Being in control Cheerfulness Hunger Receptivity Suppleness	Heaviness Sensory acuteness Tingling

Note: One item, 'Connectedness to others', showed no IIC $\geq$ 0.5, and was omitted.

This document provides background information for a conference poster displayed at the British Medical Acupuncture Society Autumn Meeting, held in co-operation with the Portuguese Medical Acupuncture Society, Friday-Sunday, 27-29 September 2013, Ipanema Park Hotel, Porto, Portugal.

Further details and related posters may be found at:
http://www.qeeg.co.uk/electroacupuncture/

Tony Steffert
tony@qeeg.co.uk

David Mayor
davidmayor@welwynacupuncture.co.uk

Research study approved by the University of Hertfordshire
Health and Emergency Professions Ethics Committee, HEPEC Protocol Number 07/11/93.
Supervisor: Professor Tim Watson, Department of Physiotherapy, University of Hertfordshire, UK.
T: 01707 284970; E: T.Watson@herts.ac.uk