The international phase 4 validation study of the EORTC QLQ-SWB32: a stand-alone measure of spiritual wellbeing (SWB) for people receiving palliative care for cancer

Authors

Bella Vivat, Teresa E Young, Julie Winstanley, Juan I Arraras, Kath Black, Frances Boyle, Anne Brédart, Anna Costantini, Jingbo Guo, M. Elisa Irarrazaval, Kunihiko Kobayashi, Renske Kruizinga, Mariana Navarro, Sepideh Omidvari, Gudrun E Rohde, Samantha Serpentini, Nigel Spry, Hanneke WM van Laarhoven, Grace M Yang. On behalf of the EORTC Quality of Life Group.

This study was funded by the EORTC Quality of Life Group.

Corresponding author: Bella Vivat: b.vivat@ucl.ac.uk

Abstract

The EORTC Quality of Life (QL) Group has just completed the final phase (field-testing and validation) of an international project to develop a stand-alone measure of spiritual wellbeing (SWB) for palliative cancer patients. Participants (n= 451) - from 14 countries on four continents; 54% female; 188 Christian, 50 Muslim, 156 with no religion - completed a provisional 36-item measure of SWB plus the EORTC QLQ-C15-PAL (PAL), then took part in a structured debriefing interview. All items showed good score distribution across response categories. We assessed scale structure using Principal Component Analysis and Rasch analysis, and explored construct validity, and convergent/divergent validity with the PAL. Twenty-two items in four scoring scales (Relationship with Self, Relationships with Others, Relationship with Someone or Something Greater, and Existential) explained 53% of the variance. The measure also includes a global SWB item and nine other items. Scores on the PAL global quality-of-life item and Emotional Functioning scale weakly-moderately correlated with scores on the global SWB item and two of the four SWB scales. This new validated 32-item SWB measure addresses a distinct aspect of quality-of-life, and is now available for use in research and clinical practice, with a role as both a measurement and an intervention tool.

Keywords

EORTC, international, measure, palliative care, questionnaire, spiritual

1. Background

There is a growing recognition that spiritual care is an important dimension of health care, particularly supportive and/or palliative care for cancer (Best et al., 2015; Phelps et al., 2012). Spiritual care is often understood as active listening to a patient's spiritual concerns (NHS Education for Scotland, 2010; Purdy, 2002; Vivat, 2008a; Vivat et al., 2013; White, 2000), but it is increasingly argued that spiritual care is not always well provided, if at all (Balboni et al., 2013; Ramondetta et al., 2013).

Tools which address spiritual issues may be useful for both initiating discussion around those issues and also assessing responses to palliative care interventions. A recent RCT of early palliative care (Zimmerman et al., 2014) used FACIT-Sp, a US-developed measure of spiritual wellbeing (SWB) (Peterman et al., 2002) as the primary outcome measure. There is, however, no "gold standard" measure of SWB, and, while there is some commonality and overlap between existing definitions of SWB and spirituality (Edwards et al., 2010), there is no generally agreed definition of either (Höcker et al., 2014; Monod et al., 2011).

Furthermore, while both functional and substantive tools for assessing and measuring spiritual issues have been developed, there are methodological concerns with some of these (Monod et al., 2011; Vivat, 2008b). Measures in this field address sometimes complex and subtle concepts, so it is important that they be developed, and not just validated, cross-culturally and in the languages where they may be used, so any linguistic and/or conceptual difficulties may be addressed at the start, rather than only during later field-testing and validation (Acquadro et al., 2012; Vivat et al., 2013; WHOQOL SRPB Group, 2006). However, no current functional tools have been developed from the outset in multiple cultural and linguistic contexts, although some have been later validated in diverse populations (e.g. Murphy et al., 2010). In contrast, some substantive tools have been developed cross-culturally (e.g. the World Health Organisation Quality of Life Group (WHOQOL) measure of Spiritual, Religious and Personal Beliefs (SRPB) (WHOQOL SRPB Group, 2006)).

The Module Development Guidelines of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG) (Johnson C. et al., 2011) recommend a cross-cultural, multi-lingual approach to developing modules, or questionnaires. Members of the EORTC QLG have developed a stand-alone functional measure of spiritual wellbeing for people receiving palliative care for cancer, following these Guidelines, in a study conducted from the outset in multiple languages and countries, initially across Europe, latterly extending beyond Europe (Vivat, 2008b; Vivat et al., 2013).

A literature review when the project began identified 84 relevant issues, grouped in three domains: relationship with self and others (29 issues), existential (24 issues) and religious (31 issues) (Vivat, 2008b). These domains paralleled key dimensions identified by other scholars (e.g. Kellehear, 2000). Study participants suggested an additional six issues. By the end of Phase 3 pilot-testing, these 90 issues were reduced to 36 items, and four scales hypothesised (Vivat et al., 2013), although the scales were not published at that time because the QLG Guidelines do not require scales to be identified at pilot-testing, and the sample size was too small for this hypothesised

structure to be definitive. This paper reports on the findings from an international, cross-cultural study using the 36-item provisional measure resulting from the pilot-testing, which was conducted in 2012-14 as Phase 4, the final validation phase of the project.

2. Methods

We broadly followed the EORTC QLG Module Development Guidelines (Johnson C. et al., 2011), and, for newly collaborating languages, the EORTC translation procedure (Dewolf et al., 2009). Ethics approval for conducting the study in England and Wales was given by the SW London Rec 4 (Surrey Borders) ethics committee (Ref: 11/LO/0692), and local and/or national ethical approval given as required in all other participating countries.
A core ethical issue was that, because completion of the measure stimulates reflection on the part of respondents, a researcher and/or care professional should always be available for subsequent discussion if required (see also Vivat et al. 2013).

Eligible participants were those aged 18 and over, with advanced and incurable disease (solid tumour or haematological malignancy). People with breast and prostate cancer were all Stage 4, all others at least Stage 3. We defined "palliative" broadly: receiving treatment or care without curative intent, i.e. with the primary objective being to improve symptoms and quality of life and/or slow disease progression. We recruited and interviewed participants in hospices (day care facilities and in-patient units), outpatient clinics, hospital wards, and at home. We collected clinical and socio-demographic data from all participants, including whether participants had religious beliefs or were involved with any spiritual movement or organisation. Participants who said they did were asked the name of their religion or spiritual movement/organisation, and whether or not they actively engaged in religious or spiritual practices.

Participants completed our provisional 36-item measure and the EORTC QLQ-C15-PAL ("PAL") (Groenvold et al., 2006): a modified, validated version of the EORTC QLQ-C30 core quality-of-life questionnaire for palliative care patients. Tabachnik and Fidell (2001) recommend ten participants per item for any measure undergoing validation using factor analysis techniques. Thirty-five of our 36 provisional items used the standard EORTC QLG four-point scale (Not at all - A little - Quite a bit - Very much), so we aimed to recruit a minimum of 400 people receiving palliative care for any cancer diagnosis: ten respondents for each of those items (10 x 35 = 350), plus an additional 50 to allow for participant dropout or excessive missing data. Response data for item 36 were analysed separately, since this was a global SWB item, for which participants scored their overall SWB on a seven-point response scale: 1 (Very poor) - 7 (Excellent), the same as the overall QL item of the EORTC QLQ-C30 and EORTC QLQ-C15-PAL, plus an additional option of 0 for "don't know/can't answer".

In Phase 3 pilot-testing (Vivat et al., 2013), some people who had stated that they had no religious or spiritual beliefs, and had never had such beliefs, were annoyed when asked to respond to further items enquiring about belief. Two items in the field-tested version of the measure were therefore re-phrased as "skip" (screening) items, enquiring (i) whether respondents believed in God or in someone or something greater than themselves, and (ii)

whether they had always done so. People who responded positively to either or both of these items were asked to complete four subsequent items: two regarding trusting and feeling connected to God/someone/something greater, and two exploring changes in beliefs. We called these participants "Believers". People who responded "Not at all" to both screening items were asked to "skip" the subsequent four items and continue to the next section of the measure.

The "skip" instructions thus produced two datasets: All study participants (A), which included some who only answered 32 of the 36 items, with a subset of "Believers" (B) who answered all 36. We did not seek to compare people who answered the skip items positively with those who did not, because for many people belief is not clear-cut but a matter of degree, and may involve various levels of uncertainty. Attempting to account for the possible variations in this, creating meaningful subgroups and sampling accordingly, would have over-complicated the study, not least by producing a geographical distortion, owing to the increasing secularisation of many post-industrial Northern European societies (Norris and Inglehart, 2011).

After completing the two measures, participants took part in structured debriefing interviews, exploring whether any items were annoying, confusing, difficult, intrusive, repetitive, or upsetting. We were also interested to investigate participants' understandings of spiritual wellbeing, while two items (living one day at a time and belief in life after death) had been previously identified as potentially ambiguous and/or as having an unclear direction. We therefore also asked what participants understood by spiritual wellbeing, living one day at a time, and life after death, and asked further whether they found their understandings of life after death and living one day at a time helpful.

We conducted a thematic analysis of the qualitative data from these interviews, and also two parallel analyses of the quantitative response data (excluding response data for the global SWB item) for datasets A and B. Figure 1 summarises the steps of the analysis, and the rationale for each step.

2.1 Item response distributions, scale structure, and model fit

The first stage in the quantitative analysis was to explore item responses, examine descriptive statistics and check for any items with range restriction in the responses, that is, where any two response categories accounted for more than 95% of all responses or any single category less than 5% of all responses (Streiner and Norman, 1995). We then investigated scale structure, using Principal Component Analysis (PCA) with Oblimin rotation (ordinal variables) to explore scale grouping and identify suitable items to form scales. Scree plots and eigenvalues greater than 1 were used to identify the optimum number of factors, with a threshold value of 0.4 for item loading coefficients.

Next, we used Rasch analysis (Pallant and Tennant, 2007; Rasch, 1960/1980; Stewart-Brown et al., 2009) to assess uni-dimensionality of the scales identified, and model fit, using RUMM2030 (Rumm Laboratory Pty Ltd, 2011). Best-fitting solutions are indicated by probabilities from the Item-Trait Interaction Chi-square test greater than

0.05, after dividing by the number of items (the Bonferroni correction (Bland and Altman, 1995)). We set desirable fit residual values, for both person and item, as mean close to zero and SD less than 1.5, and taking individual item fit residual values greater than +2.5 as indicating misfit and below -2.5 item redundancy. We assessed internal consistency using the Person Separation Index (PSI), setting 0.7 as the desirable threshold for group level analysis. We inspected threshold maps for noteworthy disordering (indicating inconsistent use of the response options), and considered modifying the response scales if this might produce a significant improvement in model fit. We compared findings from these quantitative analyses with qualitative findings from the debriefing interviews.

2.2 Internal (scale) reliability

We assessed scale reliability using Cronbach's alpha coefficients (Cronbach, 1951), with a threshold of 0.7, as appropriate for group level comparisons (Fayers and Machin, 2007).

2.3 Construct validity

We assessed construct validity by exploring any Differential Item Functioning (DIF) relating to gender or religious belief for individual items in each scale. We also compared participants grouped by gender and performance status (PS). Finally, since previous studies (Costanzo et al., 2009; Gonzalez et al., 2014; Johnson KS et al., 2011) have identified a relationship between depression and religion, spirituality, and/or SWB, we grouped and compared participants by scores on the PAL Emotional Functioning (EF) scale, which comprises two items: one for feeling depressed and one for feeling tense. For all three of these known group comparisons we examined differences in median scores using appropriate non-parametric methods (Siegel, 1956) suitable for scale scores derived by summing a set of ordinal variables.

We had provisionally planned to explore DIF and compare groups with regard to age and geographical location, if possible, but these comparisons were not a primary research aim, so were not central to the study design. In the event, the data collected did not enable this; it was not possible to produce groupings by age or geographical location which were of an adequate size and also meaningful.

2.4 Convergent and divergent validity

The PAL has been criticised for the lack of a spiritual domain (Echteld et al., 2006), so we hypothesised that responses on the PAL would show at most only weak-moderate positive correlations with responses on the provisional SWB measure. We defined correlation values of >0.5 as strong, 0.3-0.5 as moderate, 0.2-0.29 as weak, and <0.2 as negligible (Büssing et al., 2005), and examined correlations between participants' scores on each of the SWB scales, on the global SWB item, and on the PAL EF scale and global QL item. Higher scores on the SWB scales and the PAL functioning scales all indicate better functioning and/or more positive states.

2.5 Test-retest reliability

We calculated intra-class correlations for all scales, taking values of 0.7 or greater as acceptable (Fayers and Machin, 2007). We collected retest data two-three weeks after baseline from a sub-group of participants with stable disease, and investigated differences between baseline and retest scores using the Wilcoxon signed-rank test.

3. Results

We recruited 458 study participants from 14 countries: Australia, Austria, Chile, China, France, Iran, Italy, Japan, Mexico, the Netherlands, Norway, Singapore, Spain and the UK. Most completed the measure in their mother tongue, other than Singapore, where participants completed the English language version, which for some was a second language. Seven participants, from Austria (2), Japan (1), Netherlands (1), China (1) and the UK (2) were missing responses for ten or more items (more than 25% of response data). We removed these seven participants from the dataset, so leaving 451 for the analysis (Figure 2).

Participants were 54% (242) female and 46% male (208) (data missing for one Japanese participant), with mean age 59.7 years (SD=13.2); median 61, range 18-89. The largest patient groups with respect to cancer disease site were lung (118 participants; 26.2%), and breast (81; 18.0%), followed by colorectal and gynaecological (Table 1). The majority (83.4%) had metastatic disease; 16.6% (41 participants) had only locally advanced disease. The majority of patients (288; 63.9%) were receiving active anti-cancer treatment when they were recruited. Most participants were relatively well: 244 (54%) had WHO performance status (PS) scores of either 0 (fully active) or 1 (restricted in physically strenuous activity); and almost 80% had WHO PS scores 0-2 (Table 1). For 245 participants (54%), the estimated survival prognosis was six months or more.

3.1 Religious/spiritual beliefs

Table 2 indicates participants' stated religious/spiritual beliefs on the sociodemographic form, and their responses to the "skip" items on the SWB measure. In some geographic regions, participants' religious faiths varied widely within and/or between countries: 74% of Japanese participants said they had no religion, with 23% identifying as Buddhist, while Singaporean participants were mostly Christian (73%), with one Buddhist (7%) and two Muslims (14%) (data not shown).

On collecting sociodemographic data, 275 participants (61%) said that they had religious beliefs or were involved with a spiritual movement or organisation, while over a third (156; 34.6%) said they did not or were not (data missing or not disclosed for the remaining 20 respondents (4.4%)). Nevertheless, of these 156 participants, more than half (83; 54%) still indicated that they had some degree of belief in God or someone or something greater than themselves when responding to the skip items on the SWB measure. Another 83 participants responded "Not at all" to both skip items, although five of these participants had previously stated that they had a religion or were involved with a spiritual movement or organisation. Thus, distinctions between people who defined themselves as religious or spiritual and people who did not were not entirely clear-cut. However, on the basis of responses to the

skip items, the usable datasets comprised 451 for dataset A (All participants), with 368 (451-83) in dataset B ("Believers").

3.2 Completion times, help, and missing data

Completion times for the SWB measure plus the PAL were recorded for all participants other than in China (22 participants) and Austria (7 participants). For participants where data were available, the mean completion time was 14.6 min (Table 3), and under 10 min in Australia (8.7 min) and Mexico (7.9 min). In Spain the mean completion time was 32.4 min, but all the Spanish participants asked to complete the measures accompanied by the researcher, rather than independently and discussing their responses afterwards (see also Vivat et al., 2013) (Table 3).

The majority of our participants (61.5%) self-completed the measures; 19.2% required practical (instrumental) help, for example with reading or writing; and 59.4% required no help at all (including 7/21 Spanish participants) (Table 4). For 11 of the 14 participating countries, less than 20% of the participants required help with understanding items. In the remaining three countries: Iran, China, and Singapore, 60%, 41%, and 33% of respondents respectively needed help with understanding a few items. Most Chinese participants (21/22; 95.5%) completed the measures orally (Table 3).

Levels of missing data were low. No item was missing more than 5% of responses, and less than 1% of responses were missing for 27 of the 32 items answered by all participants. The highest levels of missing data for all these 32 items arose for just three: "I have felt able to forgive myself for things I have done" (4.2% missing); "I have felt able to forgive others for things they have done" (2.9%); and "I believe in life after death" (2.4%) (Figure 3). Approximately 4% of response data were missing for each of the four items answered only by "Believers", relating to God/someone/something greater.

3.3 Scale structure: PCA

The Principal Component Analysis (PCA) for both A and B datasets identified an optimum structure of five factors. We named these five scales: Relationships with Others (RO), Relationship with Self (RS), Existential (EX), Relationship with Something Greater (RSG), and Change (CH). The "Relationships with self and others" domain hypothesised at the beginning of the study (Vivat 2008b; Vivat et al., 2013) divided into two (as had also been found provisionally in Phase 3), with a few items also moving between this and the Existential domain.

The standard 29 items answered by all participants (i.e. excluding the global SWB item, the two "skip" items, and the four items which only B answered) grouped in the same factors for both A and B, although with some difference in the order of factors and items (i.e. in the percentage of variance the items explained for A and for B). Table 5 presents the PCA results with the items listed in the order for A, plus the results for B, listed in the same order, showing the similarity in the factor content between the two analyses. For comparison, the results of the PCA analysis for B alone are shown in Table 6.

The four scales RO, RS, EX, and RSG are all scoring. The items forming the fifth scale (CH) comprise two for A (all respondents), which address changes in feelings about life, and two for B ("Believers") only, which address changes in beliefs. Such changes can be either positive or negative, and determining all possible variations and directions would require a disproportionate number of items. A scale score for just these four items is therefore not meaningful. However, they were retained in the measure because they enable the collection of clinically important information. The four other items (the two skip items, plus two items on trusting and feeling connected to God or to someone or something greater) which only B answered grouped together for B in addition to the other items (for All) in the RSG scale (Tables 5 and 6).

3.4 Model fit and item deletion: Rasch analysis

Rasch analysis further refined the four multi-item scoring scales, showing that for each scale, the removal of just one item improved the fit statistics (Table 7). Participants' qualitative comments were also considered in relation to the Rasch findings. Rasch analysis showed that response data for "I live one day at a time" (RSG scale) were disordered, suggesting that it was not possible to identify a direction for this item. This was reinforced by participants' definitions of this item, which were a mixture of positive (some participants indicated that they understood the item to mean taking the most benefit they could from each day) and negative (others indicated that they felt that they could only rely on being alive for one more day). This item was therefore deleted (Table 8). Response data for "I have felt dependent on others" (EX scale) had significant misfit, so this item was also deleted Response data for two further items: "I have felt able to forgive myself for things I have done" (RO scale) and "I have worried about people important to me" (RS scale) also had significant misfit and so these items were removed from their respective scoring scales. However, both may be valuable clinically, since they prompt important reflection from respondents, and they were therefore retained, but as non-scoring items.

Rasch analysis of the B dataset indicated that the "trust" item was redundant, providing no additional information to "feeling connected." Comments from a few participants regarding repetition/irrelevance for these two items confirmed this. We therefore deleted the "trust" item (Table 8), leaving "feeling connected" as a single-item scoring scale, for B only, which we called Relationship with God (RG).

3.5 Participant comments

Of the 451 participants, 188 (41.7%) had no comments on either the measure as a whole or any of its individual items. General comments were offered by 102 participants, and most were positive, with some commenting that the measure had prompted them to think about issues which they found emotional, but that this was not necessarily a negative experience. A total of 454 specific comments regarding individual items were made by 189 participants, 18 of whom made between five and seven comments each (Table 9).

Most comments regarding specific items related to item difficulty (Table 10). Ninety-seven participants (20%) said 201 times that an item was difficult to understand; with the largest number (28/97 respondents - 29% of this group; 6% of all participants) saying this for "I feel that there is more to life than we can perceive directly." Of the 43

respondents who identified any items as confusing, the greatest number (8/43 – 19% of this group; 2% of all participants) said this about the "perceive" item, and two of those eight respondents had also said that this item was difficult. It was agreed that this item had no clinical utility either, and it was therefore deleted (Table 8).

Just 19 respondents (4%) said 32 times that an item was intrusive, with the largest number within this small group (4/19) saying this about the "forgive myself" item. Similarly small numbers of respondents commented that items were either upsetting (17 respondents) or annoying (14 respondents), with four respondents saying that "I believe in life after death" was upsetting, and another four that it was annoying (Table 10). However, other respondents considered this item very important, and it is also potentially important clinically, so was retained. Just two participants, one from the UK and one from Iran, found any items repetitive or irrelevant (Table 10), all in relation to the two items on trusting or feeling connected to God or someone/something greater. This confirmed the decision following Rasch analysis to delete the "trust in God" item (Table 8).

3.6 Item response distributions

After the four items indicated had been deleted (Table 8), 22 items remained, in four scoring scales applicable to all respondents. The response distributions for these 22 items were generally good, with low levels of missing data, and only three with less than 5% in any one response category ("Not at all") (Figure 3). Two items showed some skewing: "I have felt loved by people who are important to me" (65% Very much; 3% Not at all) and "I have felt lonely" (51% Not at all; 9% Very much) (Figure 3). Nevertheless, neither of these items showed any restriction in range (i.e. for neither did any two categories have more than 95% of responses) (Figure 3).

3.7 Internal (scale) reliability

Internal reliability was good or very good for all four scoring scales. For both A and B datasets Cronbach's alpha was 0.8 or more for both the RO and EX scales, just under 0.8 for the RSG scale, and just under 0.7 for the RS scale (Table 11).

3.8 Construct validity

No DIF was observed by gender or religious belief for any item in any scale. However, when participants were grouped by gender and their scale scores compared, female participants' median scores on the RSG scale were significantly higher than those of male participants, for both A and B datasets (for A, median score for women: 70.0 vs. 46.7 for men, $p<0.001$; for B 73.3 vs. 60.0 respectively, $p < 0.001$) (Table 12). There were no other significant differences in relation to gender. With decreasing PS, participants' median scores on the RS and EX scales also decreased significantly, but median scores for the A dataset on the RSG scale increased significantly. No significant differences relating to PS in median scores for the RO scale were found. Comparing for median scores on the PAL EF scale, participants' median scores on the RS and EX scales decreased significantly with their EF scores, as did their median RO scores, but this was only by a marginal amount for each scale.

We did not conduct DIF, nor known group comparisons, for age or geographic location, because it was not possible to construct groupings for either variable that were sufficiently large and also meaningful. When respondents were placed in three roughly equal-sized groups for age, the youngest group had a wider age range (18-55) than the middle (56-65) and oldest groups (66-89). The perceptions of an 18-year old would likely differ to a greater extent from those of a 55-year old (both in the youngest group), than would those of a 55-year old (youngest group) from a 66-year old (oldest group)). Three broad groupings obscured these differences, and more meaningful groupings were too small and/or uneven in size. With regard to geographic location, it was not possible to create groupings which did not obscure important differences with respect to ethnicity and religious faith. Participants' religious faiths varied widely within and/or between countries, so groupings by geographic region (as a proxy for ethnicity) were sometimes contradicted by groupings by religious faith, which cut across geographic/ethnic categories. In particular, "Asia" as a geographic location did not map in any simple way onto ethnicity or religious faith, but encompassed a far wider diversity than that between Northern and Southern Europe, or Central or Southern Latin America and peninsular Spain.

3.9 Convergent and divergent validity

Correlations between the PAL global quality-of-life (G-QL) item and EF scale and the four scoring scales and global SWB item (G-SWB) of the SWB measure showed that these scales and items address distinct domains, although with some associations between them (Table 13; Figure 4). For both the A and B datasets, the strongest significant correlation observed between any of the scales and items investigated was internal to the SWB measure: between the G-SWB and the RSG scale (Spearman's rho 0.56 for A and 0.62 for B). The RO and EX scales moderately correlated with G-SWB for both the A and B datasets. The fourth scale, RS, only showed a negligible correlation with G-SWB for A.

With regard to correlations with the PAL G-QL and Emotional Functioning (EF) scale: For both the A and B datasets, the EX scale of the SWB measure moderately-borderline strongly correlated with G-QL (Spearman's rho for A = 0.500; B = 0.514, $p<0.0001$ for both), and moderately with the EF scale (for A = 0.409; for B = 0.421, $p<0.0001$ for both). The RS scale moderately correlated with the EF scale (for A = 0.440, for B = 0.427, $p<0.0001$ for both), and only weakly with the global QL item. G-SWB very weakly correlated with G-QL for both the A and B datasets, and showed a negligible correlation with the EF scale. The RO scale showed significant, although negligible, correlations with both G-QL and the PAL EF scale. The RSG scale did not correlate with the EF scale nor G-QL, for neither A nor B.

Thus, the significant correlations found between the SWB scales and G-SWB were stronger than, or as strong as, any found between the SWB scales/G-SWB and the PAL EF scale/G-QL. The strongest correlations found between any SWB scales or items and the PAL scale/item were for scores on the EX scale on the SWB measure, which moderately correlated with scores on G-QL and the PAL EF scale, with similar magnitudes to correlations within the SWB measure.

3.10 Test-retest reliability

We collected retest data from 49 patients in nine countries: Chile (5), China (16), France (3), Iran (1), Italy (5), Norway (5), Singapore (5), Spain (7), and the UK (2) (Table 14). The Wilcoxon signed-rank test showed only one significant difference: the median score on the Relationship with Self scale was slightly higher at the second administration. Intra-class correlations were all at or above acceptable levels (0.70 for RO, 0.78 for RS, 0.81 for RSG and 0.90 for EX).

4. Discussion

This new stand-alone measure of SWB, the EORTC QLQ-SWB32, offers possibilities for future research and clinical practice in a field where there is as yet no "gold standard". It has been developed and validated in multiple linguistic and cultural contexts; a total of 14 countries in this final validation phase, and 18 countries across all four phases of the study. A provisional measure with 36 items has been revised to now comprise 32 items, including 22 items in four scoring scales for all respondents: Relationships with Others (RO) (six items), Relationship with Self (RS) (five items), Relationship with Something Greater (RSG) (five items), and Existential (EX) (six items), plus a single-item scoring scale: Relationship with God (RG), for people who indicate that they now believe or have previously believed in God or in someone or something greater than themselves.

These scales are each scored separately; summing scales is not appropriate, since the EORTC QLQ-SWB32, as with other EORTC measures, is a profile-based, not preference-based, measure (Costa et al., 2014). EORTC measures usually comprise several multi-item scales, plus, occasionally, single items for scoring symptoms, and a global QL item, which is scored separately. However, this EORTC SWB measure is not symptom-focused, and also has an unavoidable interventionist character, so differs from typical EORTC measures (and many other assessment/ measurement tools), in that it includes some items for their clinical utility, which are not scored, because scores for those items would not be meaningful.

Thus, this final, validated version of the measure, in addition to a global SWB item, with a seven-point response scale, equivalent to the global quality-of-life item in EORTC QLQ-C15-PAL (and EORTC QLQ-C30), also includes eight non-scoring items: (i) Two "skip" items, which determine whether participants have always had or recently developed beliefs in God or someone or something greater, which can be important information of itself, and also determines whether or not participants should respond to the single-item RG scale; and (ii) Six items with clinical value: "forgive myself", "worried about important people", and two items exploring changes in feelings about life (applicable to all respondents), plus two items exploring changes in beliefs (applicable only to people who indicate belief in God or someone or something greater). The measure has been developed to use on its own, although this study tested it alongside the PAL for validation purposes, and in order to map participants' physical statuses and other issues.

None of the patient participants in this field study experienced any difficulty with completing the SWB measure and the PAL at the same point in time, and the mean completion time for both was around 15 minutes. Most

participants were able to complete the SWB measure (less than 5% of responses missing), and few required help, although 33% of the participants in Singapore required help with understanding a few items, likely due to English being a second language for most Singaporeans (Yang et al., 2014), while 59% of participants in Iran, and 41% in China required help with understanding a few items, mostly for conceptual and/or translation reasons. Singapore, Iran, and China were collaborating for the first time in this Phase of the study, and all translation issues were addressed with assistance of the EORTC translation team. The detail of these are beyond the scope of this paper, but will be discussed in future, country-specific publications.

All items in the four scoring scales showed a good range of responses, in that they received responses in all categories, with no two response categories accounting for more than 95% of responses, nor any single category less than 5% of responses, as per criteria set by Streiner and Norman (1995) (Figure 3). Responses to two items: "feeling loved" and "feeling lonely" showed some skewing, but responses still spanned the whole range for both items, and clearly any indication that someone feels "Very lonely" or "Not at all loved" is clinically important and highly relevant. Further, responses to these items should be viewed with some caution, since social desirability bias (Krumpal, 2013) may mean that people who are feeling lonely or unloved may under-represent the strength or extent of these feelings.

Addressing spiritual wellbeing may require raising potentially challenging issues such as these, and it is therefore important that an SWB measure include related items. However, since some people may find them difficult or problematic, care should be taken when administering the measure, with support being made available after completion, if needed. We considered deleting those few items which small numbers of our participants identified as most problematic, but also took account of their possible clinical relevance before making any final decision. We deleted one item: "more to life than we can perceive directly" (identified as difficult by 28/451 participants), because it had no clinical utility, while retaining others (e.g. "forgive myself") which a few other participants (4/451) found intrusive.

When responding to the SWB measure, far fewer of our 451 participants said that they had no religious or spiritual beliefs than their responses to the sociodemographic forms indicated. In all, 156 people said when sociodemographic data were collected that they did not have a religion, nor any involvement with a spiritual movement or organisation, but just 83 responded "Not at all" to both skip items asking about belief in God or someone or something greater than themselves. This "Not at all" group also included five people who had said that they *did* have a religion or spiritual involvement when sociodemographic data were collected. Conversely, another 83 participants who had said when socio-demographic data were collected that they did *not* have any religion or spiritual involvement, still indicated in their responses to the "skip" items that they had some degree of belief in God/someone/something greater. Thus, identifying the presence or absence of religious faith and/or spiritual beliefs is not always straightforward, and may not be as easy as asking one simple question. People who state when an initial clinical history is taken that they do not follow any organised religious or spiritual grouping, may therefore

still be open to the possibility that there is something or someone greater than themselves, and this should be taken into account when planning interventions.

Despite this additional complexity, internal consistency/reliability was generally strong for the four scoring scales, including the RSG scale, which was answered by all respondents. Thus, this study indicates that the measure is suitable for people with various religious faiths, and none, and the data show that some people who indicate that they have no religious or spiritual beliefs, and have never had such beliefs, may still express belief in an afterlife in some sense, and/or value prayer, possibly as an indication of others' concern. Such people are, however, likely to score lower on the RSG scale, and are also unlikely to wish for any related intervention/s which might increase their scores. RSG scores for such people should therefore always be interpreted with caution, and with specific regard to each individual respondent.

We found no DIF for gender, nor for religious belief, but the known group comparisons for gender, PS, and PAL EF scores showed some significant differences in median scores. Women in both A and B groups scored significantly higher on the RSG scale, and this finding parallels numerous other studies which have identified that women tend to be more religious than men (Trzebiatowska and Bruce, 2012). Participants' median scores on the RS, RO and EX scales decreased significantly with decreasing scores on the PAL EF scale, similarly to findings from other studies exploring the relationship between depression and spirituality or SWB (Costanzo et al., 2009; Gonzalez et al., 2014; Johnson KS et al., 2011), although, as noted, there is no "gold standard" measure in this field, and Costanzo et al. (2009) used various measures of religious practices and experiences, while Gonzalez et al. (2014) and Johnson KS et al. (2011) both used the FACIT-Sp (Peterman et al., 2002). These measures are conceptually distinct from ours, exploring related, but not identical issues.

We did not conduct DIF analyses nor known group comparisons for age or geographic location. We were unable to construct meaningful groups for age since the only reasonably-sized groupings obscured important differences between younger and older people. Groupings by geographic region collapsed together distinct religious faiths which crossed geographic/ethnic categories, and so risked confounding variables, especially for "Asia". Findings from DIF and known group comparison analyses are important for single arm studies where patients are later divided into subgroups for comparison. However, subgroup differences would not affect multi-arm studies which match for the characteristic concerned. The details of the sometimes subtle differences between our participants are beyond the scope of the initial analysis and the current paper, and we are developing suitable hypotheses for further investigation and analysis, which will be discussed in future papers.

Correlations between responses to the QLQ-C15-PAL global QL and the SWB measure confirmed that the two measures address different, although related, dimensions of QL. Respondents' perceptions of 'overall spiritual wellbeing' and 'overall quality of life' were distinct, and scores for SWB scales and items were more closely correlated with each other than with scores for the PAL G-QL or EF scale. The largest significant correlation was within the SWB measure: scores on the RSG scale strongly correlated with scores on G-SWB. G-QL scores only

weakly correlated with G-SWB scores, and scores on the PAL EF scale and G-QL weakly or just moderately correlated with just two SWB scales. Moderate correlations between scores on the PAL EF scale, and scores on the RS and EX scales on the SWB measure reinforced the significant findings from the known group comparisons for EF scale scores. Findings from other studies comparing quality-of-life and SWB measures are similar (Bredle et al., 2011; Büssing et al, 2005), although, again, these studies used other SWB measures which are distinct from ours.

4.1 Study strengths

Participants in this validation study were highly diverse culturally and linguistically, from 14 countries and speaking ten languages; with two languages used in three countries each (English in Australia, Singapore, and the UK; Spanish in Chile, Mexico, and Spain). Multiple countries and languages were involved from the beginning of the study (Vivat et al., 2013), with more in the later phases, and the rigorous translation process ensured that concepts were meaningful and intelligible in all participating languages from the outset. In this final validation phase we recruited roughly equal proportions of men and women, and with a good age distribution, although, as typically for a palliative care population, the majority were aged over 55. Our study has kept a clear focus throughout on palliative cancer patients, so as to reduce confounding variables from other life-limiting diseases with distinct trajectories, possible complications and comorbidities, and treatments, which could impact upon SWB (Curtis et al., 2002; Galfin et al., 2010; Murray and McLoughlin, 2012).

4.2 Study limitations

We conducted our initial literature search in English language journals, and a high proportion of the literature obtained was from studies involving White Christians, many in the US, and some in non-palliative care settings (Vivat, 2008b). Nonetheless, any possible bias in initially identifying issues was addressed through the thorough development process, which explored the relevance of issues with patient and professional participants in multiple cultures and languages, including actively soliciting suggestions for additional issues. Subsequent reviews of the fast-growing body of related literature did not identify any new issues.

The study perhaps also lacks diversity within participating countries. In the UK we sought to recruit from Black and Minority Ethnic (BME) communities so that we might be able to compare responses from BME people with those from the majority community in the UK, and also with respondents of the same religion/ethnic ancestry in other countries. However, this proved difficult, and it is well understood that BME communities under-utilise palliative care services (Gunaratnam, 2006), so recruitment for palliative care studies is also challenging. Our study also lacks data from Jewish and Hindu respondents, and from Muslims in countries other than Iran, despite also actively seeking collaborators in Israel and in India.

The EORTC QLQ-SWB32 does not include all possibly relevant issues, so cannot be, nor does it claim to be, exhaustive. Spiritual wellbeing is to some extent a subjective concept, and can cover a wide range of issues, each of which will have distinct relevance for each individual. Our study began with 90 issues, which across the development and validation process have reduced to 32. It would clearly be impossible to include all possibly

relevant issues in a single measure while still keeping it manageable, especially by palliative care patients, who may be particularly prone to fatigue. It is however important to note that the measure is not only a measurement/ assessment tool, but also, unavoidably, prompts reflection and initiates discussion with patients. Respondents engage with and reflect upon the items the measure contains, and this may mean that they raise related issues, not included in the measure itself. This reflective response also means that it is ethically and clinically necessary to offer anyone who completes the measure the opportunity to discuss their responses at, or immediately after, completion. Thus, *inter alia,* it is not appropriate for respondents to complete the measure when entirely alone, e.g. at home.

Some of our study participants (including all those in Spain) indeed requested company while they were actually completing the measure, and this may have had an impact on their responses. The developers of the EORTC Elderly Care module similarly found that participants in Spain (and also some in the UK) wished to have company while completing that tool (Johnson et al., 2010). However, this did not seem to affect participants' qualitative responses on that study, and, as they point out, interviewer administration of questionnaires is valid. Our study found no differences between these individuals and other respondents who completed the measure on their own. Across all participants small increases in scores were observed for all scales on retesting, and this might be due to the research interaction. However, none of these increases were large enough to have any clinical importance, while only one difference, for scores on the RS scale, was significant.

Our definition of "palliative" was broad, and many of our participants relatively well, so our study did not include a large number of people close to the ends of their lives. People in such situations, and/or feeling less well, might have particularly pressing spiritual needs, but might find a 32-item measure challenging, and caregivers might be less willing to offer it to them. A shorter version of the measure might be preferred for such people, and a further study to investigate reducing its length would therefore be of value.

We did not explore the validity of the measure with cancer patients without palliative diagnoses, either. Future studies to examine this would also be of value, as would investigations of the validity and use of the measure with people with life-limiting conditions other than cancer, e.g. neurological conditions such as MS and MND, and with caregivers, with older people, or with people living with other chronic conditions. Comparative studies involving a greater proportion of people with religious faiths other than Christianity, and with absolute certainty of holding no religious/spiritual beliefs, would likewise be valuable.

5. Conclusions

The EORTC QLQ-SWB32 has been validated in 14 countries and ten languages, and addresses a distinct domain for quality of life. It is a stand-alone measure, comprising four scoring scales with 22 items, applicable to people with and without religious faith. It should be manageable for most palliative care populations, and studies which wish to include SWB, with potential for inclusion as an outcome measure in research studies in palliative and supportive care, particularly service evaluations and explorations of complex interventions. It would also be

valuable for before-after investigations of interventions which seek to provide spiritual care, whether explicitly, e.g. chaplaincy, Dignity therapy, or indirectly, e.g. complementary or alternative therapies such as Reiki or art therapy.

The EORTC QLQ-SWB32 and its accompanying scoring instructions are available from: groups.eortc.be/qol, as is the EORTC QLQ-C15-PAL.

Acknowledgments

We are grateful to all the patient participants and international colleagues who assisted with the study. We also thank Sheila Fisher, Eva Greimel, and Eva Nagele of the EORTC Quality of Life Group, who recruited some patients to the study, and Dagmara Kulis and colleagues in the EORTC Translation Unit for conducting and coordinating the translations.

References

Acquadro C, Conway K, Giroudet C, et al (2012). *Linguistic Validation Manual for Health Outcome Assessments*. Lyon: Mapi Research Institute. https://eprovide.mapi-trust.org/products (accessed 24 October 2016)

Balboni MJ et al. (2013). Why is spiritual care infrequent at the end of life? Spiritual care perceptions among patients, nurses, and physicians and the role of training. *J Clin Oncol*, 31: 461-7.

Best M et al. (2015). Do patients want doctors to talk about spirituality? A systematic literature review. *Patient Educ Counselling*, 98:1320-8.

Bland J, Altman D (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal*, 310: 170.

Bredle J et al. (2011). Spiritual well-being as a component of health-related quality of life: the Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being Scale (FACIT-Sp). *Religions*, 2: 77-94.

Büssing A et al (2005). Role of religion and spirituality in medical patients: Confirmatory results with the SpREUK questionnaire. *Health and Quality of Life Outcomes*, 3:10.

Costa DSJ et al. (2014). Deriving a preference-based utility measure for cancer patients from the European Organisation for the Research and Treatment of Cancer's Quality of Life Questionnaire C30: a confirmatory versus exploratory approach. *Patient Relat Outcome Meas*, 5: 119-29.

Costanzo ES et al (2009). Psychosocial adjustment among cancer survivors: findings from a national survey of health and well-being. *Health Psychol*, 28(2): 147-56.

Cronbach L (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297-334.

Curtis JR et al. (2002). Patients' perspectives on physician skill in end-of-life care: Differences between patients with COPD, cancer, and AIDS. *Chest*, 122(1):356-62.

Dewolf L et al. (2009). EORTC Quality of Life Group Translation Procedure. 3rd edition. Brussels: EORTC.

Echteld MA et al. (2006). EORTC QLQ-C15-PAL: the new standard in the assessment of health-related quality of life in advanced cancer? *Pall Med*, 20: 1-2.

Edwards A et al. (2010). The understanding of spirituality and the potential role of spiritual care in end-of-life and palliative care: a meta-study of qualitative research. *Palliat Med*, 24(8): 753-70.

Fayers PM, Machin D (2007). *Quality of Life: Assessment, Analysis and Interpretation*. Chichester, UK: Wiley.

Galfin JM et al. (2010) Psychological distress and rumination in palliative care patients and their caregivers. *J Palliat Med*, 13(11): 1345-8.

Groenvold M et al. (2006). The development of the QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *Eur J Cancer*, 42(1): 55-64.

Gonzalez P et al (2014). Spiritual well-being and depressive symptoms among cancer survivors. *Support Care Cancer*, 22(9): 2393-400.

Gunaratnam Y (2006). *Widening Access to Hospice Care: A Briefing Paper for Managers and Trustees.* London: Help the Hospices. http://www.hospiceuk.org/what-we-offer/publications?kwrd=Gunaratnam (accessed 24 October 2016).

Höcker A et al. (2014). Exploring spiritual needs and their associated factors in an urban sample of early and advanced cancer patients. *EJCC*, 23: 786

Johnson C et al., EORTC Quality of Life Group (2010). Development of the European Organisation for Research and Treatment of Cancer quality of life questionnaire module for older people with cancer: The EORTC QLQ-ELD15. *European Journal of Cancer*, 46: 2242-52.

Johnson C et al., EORTC Quality of Life Group (2011). Guidelines for developing questionnaire modules. 4th edition. Brussels: EORTC.

http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_developing_questionnaire-_final.pdf (accessed 24 October 2016).

Johnson KS et al. (2011). Which domains of spirituality are associated with anxiety and depression in patients with advanced illness? *J Gen Intern Med* 26(7): 751-8.

Kellehear A (2000). Spirituality and palliative care: a model of needs. *Palliat Med.*, 14: 149-55.

Krumpal I (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant*, 47: 2025-47.

Monod S et al. (2011). Instruments measuring spirituality in clinical research: a systematic review. *J Gen Intern Med*, 26(11): 1345-57.

Murphy PE et al. (2010). An examination of the 3-factor model and structural invariance across racial/ethnic groups for the FACIT-Sp: a report from the American Cancer Society's Study of Cancer Survivors-II (SCS-II). *Psycho-Oncology*, 19: 264–272.

Murray SA, McLoughlin P (2012). Illness trajectories and palliative care: Implications for holistic service provision for all in the last year of life. Pp. 30-51 in Sallnow L, Kumar S, Kellehear A (eds.), *International Perspectives on Public Health and Palliative Care*. London and New York: Routledge.

NHS Education for Scotland (2010). *Spiritual Care Matters: An Introductory Resource for all NHSScotland Staff.* http://www.nes.scot.nhs.uk/education-and-training/by-discipline/spiritual-care/about-spiritual-care/publications/spiritual-care-matters-an-introductory-resource-for-all-nhs-scotland-staff.aspx (accessed 24 October 2016)

Norris P, Inglehart R (2011). *Sacred and Secular: Religion and Politics Worldwide*. 2nd edition. Cambridge, New York, Melbourne: Cambridge University Press.

Pallant J, Tennant A (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale HADS. *British Journal of Clinical Psychology* 46: 1-18.

Peterman AH, Fitchett G, Brady MJ, Hernandez L, Cella D (2002). Measuring spiritual well-being in people with cancer: the functional assessment of chronic illness therapy--Spiritual Well-being Scale (FACIT-Sp). *Ann Behav Med*, 24(1):49-58.

Phelps AC et al. (2012). Addressing spirituality within the care of patients at the end of life: perspectives of patients with advanced cancer, oncologists, and oncology nurses. *J Clin Oncol*, 30: 2538-44.

Purdy, WA (2002). Spiritual discernment in palliative care. *J Palliat Med*, 5: 139–41.

Ramondetta LM et al. (2013). Surprising results regarding MASCC members' beliefs about spiritual care. *Support Care Cancer*, 21: 2991-8.

Rasch G (1960/1980 expanded edition). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen; Danish Institute for Educational Research. Expanded edition. Chicago: University of Chicago Press.

RUMM Laboratory Pty Ltd., 2011. RUMM2030 software. http://www.rummlab.com.au. Perth, Western Australia.

Siegel S (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

Stewart-Brown S et al. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes*, 7: 15.

Streiner D, Norman G (1995). *Health measurement scales: a practical guide to their development and use*. 2nd edition. London: Oxford University Press.

Tabachnick BG, Fidell LS (2001). *Using Multivariate Statistics*. 4th edition. Xxvi. Boston, Mass. and London: Allyn and Bacon.

Trzebiatowska M, Bruce S (2012). *Why are Women more Religious than Men?* Oxford: Oxford University Press.

Vivat B (2008a). "Going down" and "getting deeper": physical and metaphorical location and movement in relation to death and spirituality in a Scottish hospice. *Mortality*, 13: 42–64.

Vivat B (2008b). Spiritual issues for palliative care patients: a literature review. *Palliat Med*, 22(7): 859-68.

Vivat B et al. (2013). Cross-cultural development of the EORTC QLQ-SWB36: A stand-alone measure of spiritual wellbeing for palliative care patients with cancer. *Palliat Med*, 27(5): 457-69.

White G (2000). An inquiry into the concepts of spirituality and spiritual care. *Int J Palliat Nurs*, 6: 479-84.

WHOQOL SRPB Group (2006). A cross-cultural study of spirituality, religion, and personal beliefs as components of quality of life. *Soc Sci Med*, 62:1486–97.

Yang G, Young T, Vivat B (2014). EORTC Spiritual Wellbeing measure – preliminary results from Singapore. Abstract P059 presented at 8th EAPC Research Congress, Lleida, Spain, published in *Palliat Med*, 28(6): 538-913.

Zimmerman C et al. (2014). Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial. *Lancet*, 383: 1721-20.

LIST OF TABLES

LIST OF FIGURES

Table 1: Gender, age, disease site, metastases, symptoms and treatment, WHO performance statuses and estimated prognoses (N=451)

			frequency	%
Gender	Male		208	46.1
	Female		242	53.7
	Missing		1	0.2
Age	<= 55		149	33.0
	56–65		149	33.0
	66+		153	34.0
Disease	Non-cancer		2	0.4
	Cancer site	Lung	118	26.2
		Breast	81	18.0
		Colorectal	37	8.2
		Gynaecological	37	8.2
		GI	24	5.3
		Pancreas	24	5.3
		Prostate	22	4.9
		Haematological	13	2.9
		Head and neck	11	2.4
		Bone	10	2.2
		Melanoma	8	1.8
		Cholangiocarcinoma	7	1.6
		Bladder	6	1.3
		Liver	5	1.1
		Brain	3	0.7
Metastases[a]	None (locally advanced)		41	9.1
	Visceral		159	35.3
	Bone		154	34.1
	Liver		124	27.5
	Soft tissue		98	21.7
	Brain		62	13.7
Symptoms/ treatment	Asymptomatic/no treatment		37	8.2
	Asymptomatic/receiving treatment		78	17.3
	Symptomatic/no treatment		116	25.7
	Symptomatic/receiving treatment		210	46.6
	Missing		10	2.2

WHO PS	0: Fully active	56	12.4
	1: Restricted	188	41.7
	2: Ambulatory	116	25.7
	3: Limited self-care	55	12.2
	4: Completely disabled	30	6.7
	Missing	6	1.3
Estimated prognosis	>12 months	111	24.6
	6-12 months	134	29.7
	3-6 months	87	19.3
	<3 months	55	12.2
	Missing	64	14.2

[a] Some participants had metastases in more than one location

Table 2: Participants' stated religious beliefs against responses to the socio-demographic form and the "skip" items

	SOCIODEMOGRAPHIC RESPONSES				**"SKIP"/SCREENING RESPONSES (Q23/Q24)**	
Religion	**Religious/ Involved**	**Not religious/ Involved**	**Missing**	**Total**	**"Believer"**	**"Not at all" to both**
Christian	188	0	0	188	185	3
Muslim	50	0	0	50	50	0
Buddhist	22	0	0	22	20	2
Jewish	3	0	0	3	3	0
Sufi	1	0	0	1	1	0
Zoroastrian	1	0	0	1	1	0
Metaphysical belief	1	0	0	1	1	0
Other	1	2[a]	0	3	3	0
Not applicable	0	154	0	154	83	71
Missing	8	0	20	28	21	7
Total	275	156	20	451	368	83

[a] One of these two respondents (UK) said "I believe in Science"; the other (Norwegian) said "But I believe in God"

Table 3: Mean completion times and how completed by country (N=400)

Country	Mean completion time (mins) for both tools	Self-completion		Oral completion	
		Frequency	%	Frequency	%
AUSTRALIA	8.7	21	100.0	0	0.0
AUSTRIA	Not recorded	3	42.9	4	57.1
CHILE	15.0	30	65.2	16	34.8
CHINA	Not recorded	1	4.5	21	95.5
FRANCE	14.8	13	43.3	17	56.7
IRAN	14.5	10	22.2	35	77.8
ITALY	15.4	29	96.7	1	3.3
JAPAN	15.1	62	96.9	2	3.1
MEXICO	7.9	Not recorded	-	Not recorded	-
NETHERLANDS	12.7	42	82.4	9	17.6
NORWAY	13.1	18	54.5	15	45.5
SINGAPORE	16.7	12	80.0	3	20.0
SPAIN	32.4	0	0.0	21	100.0
UK	16.6	6	40.0	9	60.0
Total	**N/A**	**247**	**61.5**	**153**	**38.5**

Table 4: Help required and of what kind by country (N=433*)

	No help		Practical help		Help understanding		Supportive		Total
	Frequency	%	Frequency	%	Frequency	%	Frequency	%	N
AUSTRALIA	19	90.5	0	0.0	2	9.5	0	0.0	21
AUSTRIA	3	42.9	2	28.6	1	14.3	1	14.3	7
CHILE	29	61.7	12	25.5	4	8.5	2	4.3	47
CHINA	0	0.0	12	54.5	9	40.9	1	4.5	22
FRANCE	19	63.3	5	16.7	2	6.7	4	13.3	30
IRAN	15	34.1	3	6.8	27	60	0	0.0	45
ITALY	15	50.0	5	16.7	6	20.0	4	13.3	30
JAPAN	57	89.1	5	7.8	1	1.6	1	1.6	64
MEXICO	21	61.8	5	14.7	3	8.8	5	14.7	34
NETHERLANDS	41	80.4	1	2.0	6	11.8	1	2.0	49
NORWAY	18	56.3	12	37.5	1	3.1	1	3.1	32
SINGAPORE	10	66.7	0	0.0	5	33.3	0	0.0	15
SPAIN	7	33.3	11	52.4	1	4.8	2	9.5	21
UK	3	20.0	10	66.7	2	13.3	0	0.0	15
Total	**257**	**59.4**	**83**	**19.2**	**70**	**15.9**	**22**	**5.1**	**432**[a]

[a] All but two of the participants who required help required only one kind of help. The other two, both from the Netherlands, needed both practical and supportive help, with one also needing help with understanding.

Table 5: PCA analysis for All (A), with values for Believers (B) added for comparison

	RO (A)	RO (B)	RS (A)	RS (B)	RSG (A)	RSG (B)	EX (A)	EX (B)	CH (A)	CH (B)
(11) Able to trust others	0.842	0.852								
(8) Loved by those important to me	0.714	0.624								
(9) Someone to talk to about my feelings	0.708	0.664								
(12) Able to forgive others	0.681	0.667								
(13) Valued as a person	0.651	0.587								
(7) Share thoughts with those close to me	0.625	0.593								
(3) Forgive myself	0.513	0.596								
(20) Worries/concerns about the future			0.72	0.757						
(5) Worried about people important to me			0.689	0.694						
(4) Troubled			0.633	0.564						
(6) Lonely			0.5	0.459						
(15) Can anything be done for me			0.477	0.572						
(16) Unfair that I am ill			0.456	0.55						
(31) I believe in life after death					-0.824	0.801				
(32) More to life than perceive directly					-0.807	0.632				
(33) Live on through words, deeds.....					-0.675	0.58				
(22) Important others pray for me					-0.667	0.583				
(35) I have spiritual wellbeing					-0.588	0.624				
(21) Time for quietness/prayer/meditation					-0.562	0.49				
(34) I live one day at a time		-0.14		-0.214	-0.434	0.233				
(28) Trust in God or someone/ something greater					X	*0.905*				
(23) Believe in God or someone/something greater					X	*0.877*				
(24) Always believed in God or someone/thing…					X	*0.876*				
(27) Connected to God or someone/thing…					X	*0.842*				
(14) Find things I enjoy							0.656	0.59		
(1) Able to deal with problems							0.588	0.531		
(10) Dependent on others							-0.581	-0.453		
(18) My life is worthwhile							0.568	0.598		
(17) My life is fulfilling							0.563	0.643		
(19) Plan for the future							0.534	0.635		
(2) Peace with myself							0.504	0.482		
(29) Feelings about life changed since felt less well									0.887	0.644
(30) Feelings about life changed in last few weeks									0.817	0.57
(26) Beliefs changed in last few weeks									*X*	*0.855*
(25) Beliefs changed since felt less well									*X*	*0.852*

A: all; B: "Believers" subset. RO: Relationships with Others; RS: Relationship with Self; RSG: Relationship with Someone or Something Greater; EX: Existential; CH: Changes.
Extraction Method: Principal Component Analysis, Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 10 iterations.
Items in italics are those which only Believers answered.

Table 6: PCA analysis for Believers (B) only

	Component				
	RSG	**RS**	**RO**	**CH**	**EX**
(28) Trust in God or someone/thing greater	*.905*				
(23) Believe in God or someone/thing....	*.877*				
(24) Always believed in God or someone/thing....	*.876*				
(27) Connected to God or someone/thing greater	.842				
(31) I believe in life after death	.801				
(32) More to life than perceive directly	.632				
(35) I have spiritual wellbeing	.624				
(22) Important others pray for me	.583				
(33) Live on through words, deeds.....	.580				
(21) Time for quietness/prayer/meditation	.490				
(34) I live one day at a time	.233	-.214	-.140		
(20) Worries/concerns about the future		.757			
(5) Worried about people important to me		.694			
(15) Can anything be done for me		.572			
(4) Troubled		.564			
(16) Unfair that I am ill		.550			
(6) Lonely		.459			
(11) Able to trust others			.852		
(12) Able to forgive others			.667		
(9) Someone to talk to about my feelings			.664		
(8) Loved by those important to me			.624		
(3) Forgive myself			.596		
(7) Share thoughts with those close to me			.593		
(13) Valued as a person			.587		
(26) Beliefs changed in last few weeks				*.855*	
(25) Beliefs changed since felt less well				*.852*	
(29) Feelings about life changed since felt less well				.644	
(30) Feelings about life changed in last few weeks				.570	
(17) My life is fulfilling					.643
(19) Plan for the future					.635
(18) My life is worthwhile					.598
(14) Find things I enjoy					.590
(1) Able to deal with problems					.531
(2) Peace with myself					.482
(10) Dependent on others					-.453

B: "Believers" subset. RO: Relationships with Others; RS: Relationship with Self; RSG: Relationship with Someone or Something Greater; EX: Existential; CH: Changes
Extraction Method: Principal Component Analysis, Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 10 iterations.
Items in italics are those which only Believers answered.

Table 7: Overall fit, Item Fit, and Person Fit statistics

		Overall Fit			Item Fit		Person Fit			PSI	% sig. t-tests <5%
SWB scales	**Analysis [No. of items]**	**X^2**	**df**	**p**	**Mean**	**SD**	**Mean**	**SD**	**Sample size**		
All participants dataset											
Relationships with Others	Original Scale **[7]**	64.244	35	0.002	0.212	1.612	-0.382	1.244	403	0.744	
Item removed: Forgive myself[a]	**Final scale [6]**	**44.063**	**30**	**0.047**	**0.093**	**1.422**	**-0.423**	**1.235**	390	**0.717**	**1.11%**
Relationship with Self	Original Scale [6]	89.015	36	<0.0005	0.497	1.434	-0.265	1.079	440	0.684	
Item removed: Worried about people important to me [a]	**Final scale [5]**	**45.387**	**30**	**0.036**	**0.574**	**1.050**	**-0.305**	**-1.028**	433	**0.604**	**0.44%**
Relationship with Something Greater	Original Scale [6]	99.62	36	<0.0005	0.594	2.277	-0.277	1.261	421	0.677	
Item removed: I live one day at a time[b]	**Final scale [5]**	**35.8**	**25**	**0.075**	**0.571**	**1.205**	**-0.294**	**1.208**	**396**	**0.66**	<0.01%-
Existential	Original Scale [7]	479.67	42	<0.0005	0.499	5.226	-0.296	1.392	445		
Item removed: Dependent on others [b]	**Final scale [6]**	**49.530**	**36**	**0.066**	**0.428**	**1.211**	**-0.373**	**1.241**	420	**0.754**	3.10%
Believers only											
Relationships with Others	Original Scale [7]	45.901	35	0.103	0.155	1.083	-0.382	1.239	311	0.745	2.01%
	Final scale [6]	**41.53**	**30**	**0.078**	**0.109**	**1.202**	**-0.427**	**1.257**	309	**0.709**	**2.59%**
Relationship with Self	Original Scale [6]	65.302	30	0.0002	0.501	1.239	-0.258	1.07	338	0.684	
	Final scale [5]	**35.486**	**25**	**0.079**	**0.595**	**0.876**	**-0.308**	**1.184**	333	**0.618**	**0.57%**
Relationship with Something Greater	**Original scale [6]**	45.222	36	0.139	0.550	1.211	-0.334	1.295	306	0.678	<0.01%
	Final scale [5]	**36.557**	**30**	**0.190**	**0.677**	**0.634**	**-0.314**	**1.259**	302	0.611	**<0.01%**
Existential	Original Scale [7]	351.93	42	<0.0005	0.426	4.385	-0.290	1.348	343		
	Final scale [6]	**25.634**	**24**	**0.371**	**0.413**	**1.314**	**-0.353**	**1.204**	322	**0.717**	**3.74%**

Threshold value calculated using Bonferroni (BF) correction: 0.05 divided by (number of items x number of test groups)
Bolded data indicates best-fitting scale
[a] Removed from scale, but retained in measure due to possible clinical relevance
[b] Removed from scale and deleted from measure – no clinical relevance

Table 8: Deleted items and reasons for deletions

Item	Rasch analysis	Definition data	Qualitative data
Trust in God or someone or something greater	X		
Live one day at a time	X	X	
Dependent on others	X		
More to life than perceive directly			X

Table 9: Socio demographic details for the 18 participants who commented on 5-7 items each

Country	Male (Age, Religion)	Female (Age, Religion)
China	1 (69, None)	-
France	-	1 (50, None) 1 (59, Christian)
Iran	1 (43, Muslim)	1 (20, Muslim)
Netherlands	1 (66, None) 1 (56, None) 1 (58, None)	1 (64, None)
Norway	-	1 (67, None) 1 (64, Christian)
Singapore	-	1(63, Christian) 1 (64, Christian)
Spain	1 (56, Christian) 1 (76, Christian)	1 (59, Christian)
UK	1 (26, "I believe in science")	1 (89, Christian)
Total	8	10

Table 10: Numbers of respondents and responses regarding problems with items, and items with most comments

	Total no. of respondents	Total no. of responses	Items with most comments
Difficult	97	201	"I feel there is more to life than we can perceive directly" (28 respondents; 9 from Iran, 6 from Spain)
Confusing	43	60	"I feel there is more to life than we can perceive directly" (8 respondents; 2 also commented the item was difficult: 1 from Chile, 1 from Mexico)
Intrusive	19	32	"I have felt able to forgive myself for things I have done" (4 respondents)
Upsetting	17	23	"I believe in life after death" (4 respondents)
Annoying	14	19	"I believe in life after death" (4 respondents, distinct from those who commented that this item was upsetting)
Irrelevant	1	3	"I feel connected to God/someone/something…" and "I trust in God/…" One UK respondent: "They are the same as I believe in God…/I have always believed in God…" (also see Repetitive) "I have able to forgive others for things they have done." The same UK respondent: "I don't think anyone's done anything wrong to me"
Repetitive	2	4	"I feel connected to God/someone/something…" and "I trust in God/…" The same UK respondent as Irrelevant. "I believe in God/someone/something…" and "I trust in God/…" One Iranian respondent (in Farsi, belief and trust differ by only one letter, so respondent misunderstood on first reading).

Table 11: Cronbach's alphas for scales for All (A) and Believers Only (B)

Scale	**Number of items**	**Cronbach's alphas for 'A' (n=451)**	**Cronbach's alphas for 'B' (n=368)**
Relationships with Others	6	0.84 (n=433)	0.84 (n=355)
Relationship with Self	5	0.68 (n=445)	0.69 (n=363)
Relationship with Something Greater	5	0.78 (n=428)	0.77 (n=351)
Existential	6	0.81 (n=440)	0.80 (n=361)

Table 12: Known group comparisons

		All RS Mean (SD) Median (Range)	All RO Mean (SD) Median (Range)	All EX Mean (SD) Median (Range)	All RSG Mean (SD) Median (Range)	Believers only RSG Mean (SD) Median (Range)
Gender	M	61.5 (21.7) 66.7 (0-100) N=204	71.1 (22.5) 72.2 (0-100) N=202	60.7 (24.6) 61.1 (0-100) N=205	52.4 (26.1) 46.7 (0-100) N=199	58.2 (25.3) 60.0 (6.7-100) N=153
	F	57.4 (23.4) 60.0 (0-100) N=240	73.5 (21.0) 77.8 (5.6-100) N=230	61.8 (22.1) 63.9 (0-100) N=234	66.36 (25.6) 70.0 (0-100) N=228	71.5 (22.4) 73.3 (13.3-100) N=198
	Mann Whitney	P=0.072	P=0.346	P=0.676	P<0.001	P=<0.001
EF	Good EF (100 – 76)	68.6 (19.9) 73.3 (0-100) N=195	75.8 (20.8) 80.6 (0-100) N=184	70.5 (22.7) 72.2 (0-100) N=190	57.8 (28.3) 60.0 (0-100) N=186	64.6 (26.6) 66.7 (6.7-100) N=145
	Moderate (75-26)	54.7 (21.8) 53.3 (0-100) N=205	69.7 (21.3) 72.2 (5.6-100) N=205	55.8 (20.3) 55.6 (5.6-100) N=206	61.5 (25.7) 60.0 (0-100) N=199	66.5 (23.3) 66.7 (6.7-100) N=171
	Poor (=<25)	40.5(20.2) 40 (0-93.3) N=43	70.1 (26.3) 75.0 (5.6-100) N= 42	45.9 (23.6) 50/0 (0-100) N=42	60.7 (25.2) 53.3 (0-100) N=41	67.1 (22.5) 73.3 (20-100) N=33
	Kruskal Wallis	P<0.001 (2DF)	P=0.010 (2DF)	P=<0.001	P=0.474(2DF)	P=0.904 (2DF)
PS	Fully active	63.8 (21.6) 66.7 (0-100) N=56	71.6 (24.9) 77.8 (16.7-100) N=51	69.7 (23.4) 72.2 (0-100) N=55	55.2 (29.5) 53.3 (0-100) N=53	61.2 (28.0) 60.0 (13.3-100) N=44
	Restricted	61.5 (22.9) 58.2 (0-100) N=186	70.7 (21.9) 72.2 (0-100) N=183	63.3 (23.2) 66.7 (0-100) N=185	56.3 (27.1) 53.3 (0-100) N=180	63.0 (24.9) 60.0 (6.7-100) N=145
	Ambulatory>50 % of the time	56.5 (21.1) 60.0 (0-100) N=113	74.8 (20.2) 77.8 (22.2-100) N=113	59.1 (23.3) 61.1 (11.1-100) N=113	63.1 (25.8) 66.7 (0-100) N=111	67.0 (24.4) 73.3 (13.3-100) N=96
	Limited	53.8 (23.0) 53.3 (13.3-93.3) N=54	73.4 (21.0) 77.8 (22.2-100) N=51	53.1 (21.4) 55.6 (5.6-94.4) N=51	65.6 (24.7) 73.3 (0-100) N=48	72.0 (19.9) 73.3 (20-100) N=40
	Completely disabled	55.1 (26.1) 53.3 (6.7-100) N=30	72.4 (21.3) 72.2 (5.6-100) N=29	54.1 (20.4) 55.6 (5.6-88.9) N=30	70.7 (21.0) 73.3 (33.3-100) N=30	76.4 (18.1) 80 (40-100) N=24
	Kruskal Wallis	P=0.033(4 DF)	P=0.669 (4DF)	P=0.001 (4DF)	P=0.011 (4DF)	P=0.048 (4DF)

Table 13: Correlations between PAL G-QL and EF scale and G-SWB and all SWB scoring scales for All participants and Believers only

		G-SWB	RSG	RO	RS	EX
A: All participants						
PAL G-QL	Correlation coefficient	.254**	.024	.162**	.271**	.500**
	Sig. (2-tailed)	.000	.615	.001	.000	.000
	N	397	428	433	445	440
PAL EF	Correlation coefficient	.162**	-.075	.128**	.440**	.409**
	Sig. (2-tailed)	.001	.129	.008	.000	.000
	N	396	426	431	443	438
G-SWB	Correlation coefficient	1.000	.562**	.343**	.118*	.444**
	Sig. (2-tailed)	-	.000	.000	.019	.000
	N	398	389	388	394	393
B: Believers						
PAL G-QL	Correlation coefficient	.216**	.034	.178**	.264**	.514**
	Sig. (2-tailed)	.000	.527	.001	.000	.000
	N	338	351	355	363	361
PAL EF	Correlation coefficient	.119*	-.051	.124*	.427**	.421**
	Sig. (2-tailed)	.029	.338	.020	.000	.000
	N	337	349	353	361	359
G-SWB	Correlation coefficient	1.000	.622**	.376**	.081	.442**
	Sig. (2-tailed)	-	.000	.000	.138	.000
	N	339	330	331	335	335

Table 14: Test-retest scale scores and Wilcoxon signed ranks test

	Mean	SD	Median	Valid N	Z (Wilcoxon) (negative ranks)	Asymp. Sig. (2-tailed)
RO	75.00	21.89	77.78	48		
RO2 (RO after 2-3 weeks)	79.17	21.02	83.33	48		
RO2-RO	4.17		5.55		-1.249[b]	.212
RS	60.95	23.77	60.00	49		
RS2 (RS after 2-3 weeks)	67.64	21.13	66.67	48		
RS2 - RS	6.69		6.67		-2.75	.030
RSG	61.67	27.52	60.00	48		
RSG2 (RSG after 2-3 weeks)	66.53	29.86	66.67	49		
RSG2 - RSG	4.86		6.67		-1.784	.074
EX	62.81	26.04	66.67	49		
EX2 (EX after 2-3 weeks)	65.31	24.10	72.22	49		
EX2 - EX	2.50		5.55		-1.181	.238

Figure 1: Steps in analysis plus rationale for each step

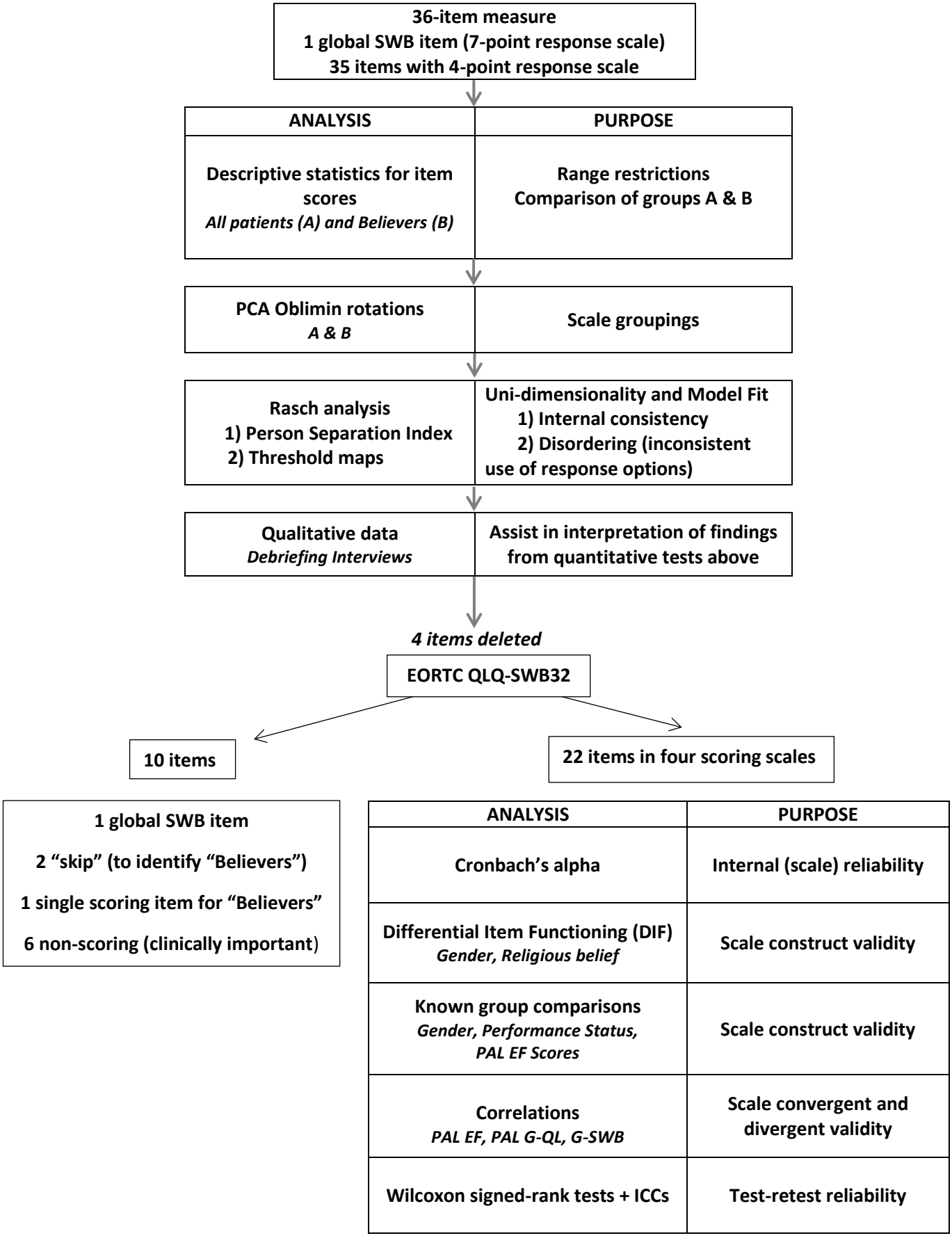

Figure 2: Numbers of participants from each country included in the analysis (N=451)

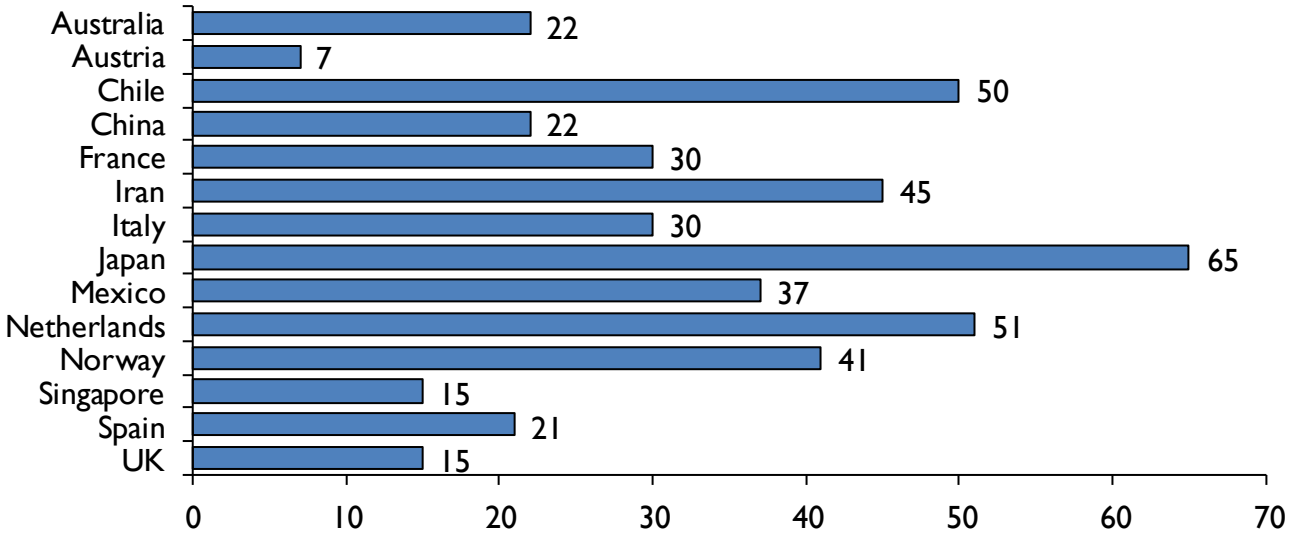

Figure 3: Response distributions for scoring items in their scales

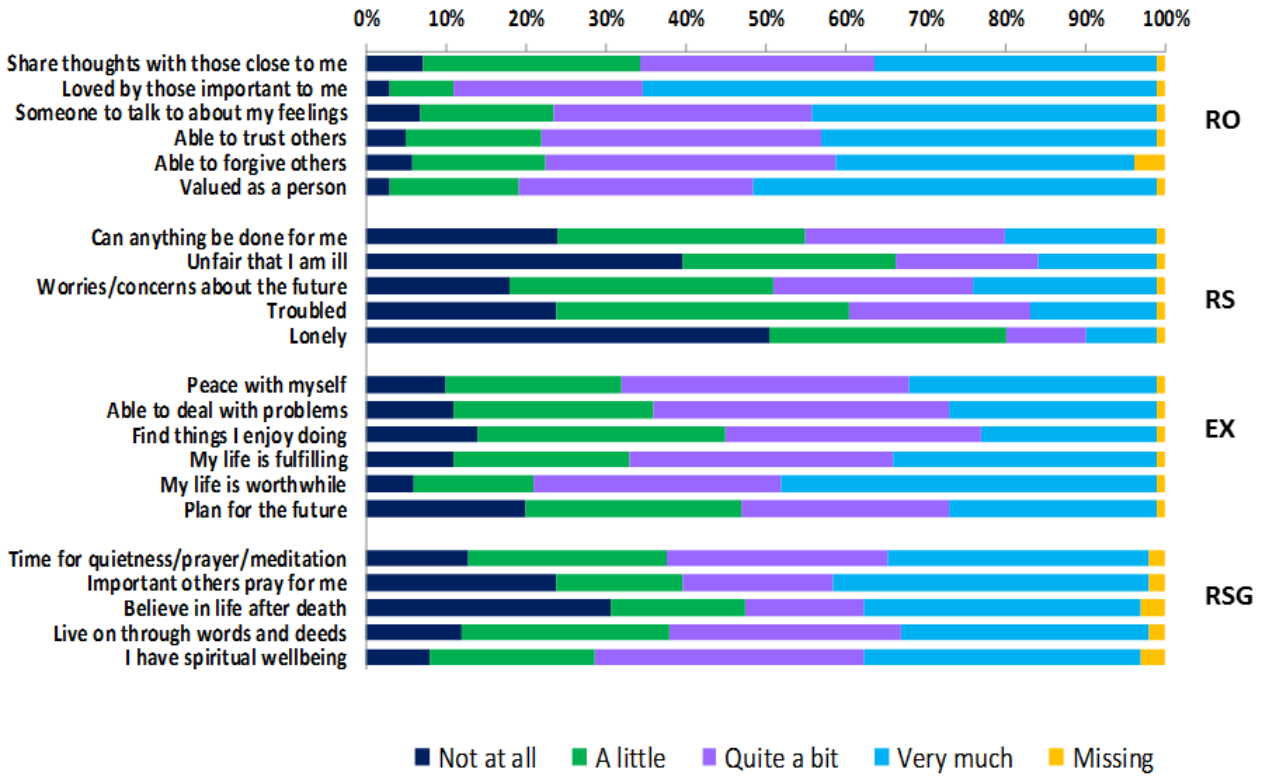

Figure 4: Correlations between SWB scales and item and PAL QL item and EF scale for All participants (A)