

1 Developing and testing a measure of consultation-based reassurance for people with low back pain
2 in primary care: A cross-sectional study

3 Nicola Holt^{1,2,*} & Tamar Pincus^{1,3}

4 ¹ Royal Holloway, University of London, Egham Hill, Egham, TW20 0EX

5 ² ztjt128@live.rhul.ac.uk

6 ³ t.pincus@rhul.ac.uk

7 * Corresponding author

8

9

10 **Abstract**

11 **Background**

12 Reassurance from physicians is commonly recommended in guidelines for the management of low
13 back pain (LBP), but the process of reassurance and its impact on patients is poorly researched.

14 We aimed to develop a valid and reliable measure of the process of reassurance during LBP
15 consultations.

16 **Methods**

17 Items representing the data-gathering stage of the consultation and affective and cognitive
18 reassurance were generated from literature on physician-patient communication and piloted with
19 expert researchers and physicians, a Patient and Public Involvement group, and LBP patients to form
20 a questionnaire. Patients presenting for LBP at 43 General Practice surgeries were sent the
21 questionnaire. The questionnaire was analysed with Rasch modelling, using two samples from the
22 same population of recent LBP consultations: the first (n=157, follow-up n=84) for exploratory
23 analysis and the second (n=162, follow-up n=74) for confirmatory testing. Responses to the
24 questionnaire were compared with responses to satisfaction and enablement scales to assess the
25 external validity of the items, and participants completed the questionnaire again one-week later to
26 assess test-retest reliability.

27 **Results**

28 The questionnaire was separated into four subscales: data-gathering, relationship-building, generic
29 reassurance, and cognitive reassurance, each containing three items. All subscales showed good
30 validity within the Rasch models, and good reliability based on person- and item-separations and
31 test-retest reliability. All four subscales were significantly positively correlated with satisfaction and
32 enablement for both samples. The final version of the questionnaire is presented here.

33 **Conclusions**

34 Overall, the measure has demonstrated a good level of validity and generally acceptable reliability.
35 This is the first measure to focus specifically on reassurance for LBP in primary care settings, and will
36 enable researchers to further understanding of what is reassuring within the context of low back
37 pain consultations, and how outcomes are affected by different types of reassurance. Additionally,
38 the measure may provide a useful training and audit tool for physicians. The new measure requires
39 testing in prospective cohorts, and would benefit from further validation against ethnographic
40 observation of consultations in real time.

41 **Background**

42 Delivering effective reassurance to people presenting with musculoskeletal, or non-specific low back
43 pain (LBP) is recommended by most guidelines, to convey the message that LBP has a good
44 prognosis, there is no need for x-rays, there is no underlying serious pathology, and patients should
45 stay active [1]. These messages are considered to enhance patients' ability to self-manage and
46 reduce long term disability. Evidence on effective reassurance in LBP remains scarce. A systematic
47 review [2] of prospective cohorts in primary care that measured practitioners' behaviours during the
48 consultation and their association with patient outcomes found only one study in LBP [3]. The
49 majority of studies included mixed groups of consecutive consultations. The findings from the review
50 suggest that while cognitive reassurance (explaining the aetiology and prognosis and discussing
51 interventions) is associated with better outcomes in primary care, affective reassurance (rapport
52 building, indications of empathy and generic reassuring statements) might improve patient
53 satisfaction, but might result in higher symptom burden later on for patients with non-specific
54 conditions. The authors refer to earlier theoretical work [4] that argues that affective reassurance
55 results in immediate reduction of anxiety, but this in turn leads to reduction in patients' engagement
56 with cognitive reassurance, breeds dependence on the practitioner, and ultimately results in worse
57 outcomes in the long run. As a result, reassurance of any kind may be expected to increase patients'
58 immediate satisfaction and enablement, as they leave the consultation still experiencing the
59 beneficial effects of the practitioner telling them that they are going to be fine, but if effective
60 cognitive reassurance has not been properly engaged with, anxiety will recur in the face of ongoing
61 symptoms. Findings from Interviews with low back pain patients [5] supported these conclusions, as
62 they describe patients' perceptions that only explicit reassurance through explanations about their
63 problem reduced participants' concerns. The participants in this sample noticed, appreciated, and
64 remembered affective behaviours and wanted to feel that their physician understood them and was
65 taking them seriously, but valued information which would help them to manage their problem
66 more highly.

67

68 The impact of physicians' consultation-based reassurance in LBP warrants further investigation. Even
69 in groups conceptualised as low-risk of long-term pain (those who do not exhibit psychological
70 obstacles to recovery) interventions are not optimal. For example, evidence from a large randomised
71 controlled trial that screened patients for risk, and offered those at low-risk minimal intervention [6],
72 based mainly on education shows that at 4 months 27% had not recovered, and 37% had not
73 recovered at 12 months. These findings suggest that for this group interventions can be improved,

74 but this requires better understanding of patients' needs, and better evidence to develop more
75 effective minimal interventions.

76

77 In order to study how consultation-based reassurance impacts on outcomes in LBP, ultimately
78 leading to improved consultations, there is a need to develop a measure of the process. Any
79 measure must be tested in relevant populations (in this case LBP patients) and demonstrate good
80 levels of reliability and validity, in order to be considered an acceptable tool for capturing
81 reassurance. There are a number of instruments designed to measure the content of consultations
82 in primary care, but none focused on reassurance, or on LBP. The aims of this study were:

- 83 1. To develop and test a theory-driven reliable and valid questionnaire to assess consultation-
84 related reassurance in LBP, and
- 85 2. The subsequent selection of a short version by removing similar items to ensure our final
86 instrument is easily usable.

87 **Methods**

88 **Generation of items:**

89 For the purposes of this review Linton et al.'s [7] definition of reassurance was used:

90

91 "reassurance '...removes the fears or doubts of (pain/illness); to comfort'. Reassurance
92 always takes place within the dynamics of the interaction between the caregiver who has
93 the intention to reduce worry, and the patient who is concerned. Ultimately, reassurance is
94 achieved if the patient changes his/her behavior, understanding or thoughts." [7, pp. 5]

95

96 Therefore, reassurance was defined as any behaviour by a physician which could lead to reduced
97 worry in a concerned patient, and further classified according to the model of reassurance
98 developed by Pincus et al [2]. In the first instance, specific examples of physicians' behaviours during
99 consultations were extracted from the literature. We identified theoretical reviews and empirical
100 studies of patient-centred consultation to provide a comprehensive description of the variety of
101 behaviours associated with reassurance. From these reviews, physician behaviours which were
102 theoretically or evidentially associated with improved outcomes post-consultation were extracted.
103 Classification of the identified behaviours according to the model [2] allowed for the formulation of
104 conceptual maps describing different aspects of the consultation. The model describes 3 global
105 concepts: At earlier stages of the consultation, data-gathering included demonstrating
106 understanding of the patient's problem; eliciting patients' concerns and finding out the whole story
107 (see figure 1). At later stages of the consultation, cognitive reassurance (see figure 2) includes giving
108 information about aetiology, prognosis and treatment options; giving patients a chance to ask
109 questions; checking that patients understand the information and the recommendations and
110 matching the information to individual patient concerns and whole story. The final concept (see
111 figure 3), Affective Reassurance, includes giving generic reassurance; showing confidence; giving a
112 clear message that uncertainty (in reference to cause/aetiology of the problem, prognosis and/or
113 response to treatment) is manageable; showing care and empathy and building a relationship with
114 the patient.

115

116 From these conceptual maps, items were generated under each of the three headings. The items
117 were sent out to a team of expert low back pain researchers, including a psychologist, an osteopath,

118 and two General Practitioners (GPs) for comments. This feedback was used to modify the item pool,
119 change wording where required and add or remove items as recommended. The final pool of items
120 consisted of 30 items: 7 data-gathering; 9 cognitive reassurance; and 14 affective reassurance. The
121 items on data gathering appeared first, followed by the items on cognitive and affective reassurance,
122 which randomised. The questions were preceded by the instructions: 'To what extent did the
123 physician', and the response mode was a 7 point Likert scale, with the anchors ranging from 'not at
124 all' to 'a great deal'.

125

126 Advice on the questionnaire was sought from a Patient and Public Involvement (PPI) group based in
127 Surrey, UK, who indicated that the items were acceptable and understandable. They recommended
128 minor changes in wording, which were applied to the questionnaire. Participants in another study [5]
129 also agreed to read and comment on the questionnaire. Again, the consensus was positive on the
130 item content and presentation.

131

132 **Testing of the new questionnaire**

133 Participants

134 Forty-three general practice surgeries in the UK recruited patients presenting for a new episode of
135 LBP between October 2013 and April 2015. Patients were identified by a database search using a
136 search strategy developed specifically for the study by an independent expert company (Holt et al.,
137 2015). The searches were carried out once a month by each practice. The searches were conducted
138 by a researcher at the practice (such as a designated research nurse), and were checked by GPs to
139 ensure that identified patients were eligible and suitable to participate. The practice then sent out a
140 study pack to eligible patients containing the documents outlined below.

141 The inclusion and exclusion criteria used to identify eligible patients were as follows:

142

143 Inclusions:

144 Consultation within the previous month.

145 New episode of acute LBP (duration <6 weeks; no prior episodes within last 6 months) without
146 radiating leg pain and for whom self-management was indicated (i.e. those not offered follow-up
147 care).

148 Adult patients (>18 years).

149 Exclusions:

150 Red flag markers.

151 Cancer.

152 Cauda equina and ankylosing spondylitis.

153 Severe disability or end of life disorder.

154 Pregnancy.

155 Cognitive impairment or serious mental health problems, which the GP considers could make
156 patients vulnerable and for whom participation would be detrimental.

157 Previous spinal surgery.

158 Currently receiving secondary care (physiotherapy, osteopathy, etc.) for the same problem.

159 Unable to read and speak English.

160 Those requiring further investigation.

161

162 **Materials and Procedures**

163 The Questionnaire packs sent to participants contained: a letter of invitation; a study information
164 sheet; a consent form; a questionnaire; and a form to opt in to complete the reassurance
165 questionnaire a second time, one week later, for the purposes of temporal (test-retest) reliability
166 analysis. The following information was collected at the same time as participants' initial responses
167 to the questionnaire:

168 Demographic Information

- 169 • Age
- 170 • Gender
- 171 • Physician gender
- 172 • Type of physician (GP or nurse)
- 173 • Marital status
- 174 • Education level
- 175 • Employment status

176 Pain and Function

- 177 • Length of current episode of LBP
- 178 • Whether or not this is the participant's first episode of LBP
- 179 • Number of previous GP consultations for this episode
- 180 • Details of any other physician participants had seen since their consultation
- 181 • Pain intensity in the week prior to their consultation, rated on the 11-point Pain Numeric
- 182 Rating Scale ranging from 0 (no pain) to 10 (worst possible pain) [NRS, 8].
- 183 • Functional status was assessed using the Roland-Morris Disability Questionnaire [RMDQ, 9]
- 184 which is a well-validated measure of disability in low back pain populations [10].

185 Consultation outcomes

- 186 • To measure satisfaction, the Consultation Satisfaction Questionnaire [CSQ, 11] was used.
- 187 The CSQ is a validated 9-item questionnaire in which participants respond to statements
- 188 about how they felt about the consultation on a five-point scale from 'strongly agree' to
- 189 'strongly disagree'.
- 190 • Enablement was measured with the Patient Enablement Instrument [PEI, 12] which has
- 191 been validated for use in primary care populations [13]. The PEI consists of 6 items, rated on
- 192 a 3-point scale from either 'much better' to 'same or less' or 'much more' to 'same or less'.

193

194

195 **Analysis**

196 Item-Response Theory

197 Item Response Theory (IRT), originally developed in educational settings, has grown in popularity
198 within the psychological and health sciences in recent years for constructing measures [e.g. 14, 15,
199 16]. IRT is based on item response functions, which are mathematical functions describing the
200 relationship between a person's probable response to a scale item and where he/she falls on the
201 continuum of the construct being measured by that item [15, 16]. IRT models aim to construct
202 measures which accurately assess latent (unobservable) traits, and it is assumed that a person must
203 have a higher level of the trait to score highly on more difficult items. IRT models were originally
204 developed for dichotomous items, but have been extended to include items with nominal response
205 options, such as Likert scales.

206 The mathematical models used within IRT are independent of sample data, and so comparison of
207 responses across groups becomes possible [17]. Additionally, each item is scrutinised, to reduce
208 redundancy as well as ensuring that the scale is valid and reliable. One of the most commonly used
209 IRT models is the Rasch Measurement Model [18-20], which is used in this analysis. Rasch analysis
210 allows for validity and reliability testing within the same model, and accounts for missing data by
211 using the expected scores (for a person's ability on a question's difficulty level) where no score has
212 been given. In this analysis the one-parameter Rasch rating scale model (RSM) is used, which is an
213 extension of the simple (dichotomous) Rasch model for rating scale observations like the present
214 one. The model allows the item difficulty (in this case the extent to which each behaviour is reported
215 to have been present) to be based on the way in which an appropriate group of subjects (i.e. the
216 patients) actually responded to that question, and establishes the relative difficulty of each item
217 stem in recording the development of an attitude from the lowest to the highest levels the
218 instrument is able to record, i.e. from response categories 1 to 7 [21, 22].

219

220 This study employed a cross-sectional design; all data were taken from participants at a single time-
221 point, with the exception of the reassurance questionnaire which was answered for a second time
222 one week after the first in order to assess test-retest reliability. Two separate samples were
223 obtained for this study: the first 150 participants, referred to as Sample 1, for an exploratory analysis
224 of the questionnaire; the second 150 participants (Sample 2) were new participants recruited from
225 the same pool of practices for confirmatory testing. Potential participants who had already been
226 invited to take part in the study had a study-specific Read code entered into their notes, which
227 allowed us to exclude those already invited from future searches, should they have consulted again
228 within the study period. All analyses were conducted on both samples, with the exception of
229 Dimensionality Mapping (see 'Structural Validity, below), which identified subscales within the
230 questionnaire from Sample 1's data only. See Figure 4 for a representation of the collection and
231 analysis of data for this study. Analyses were conducted using Winsteps version 3.8.1.0 computer
232 software [23] and following the guidance for conducting and reporting Rasch analysis set out by
233 Tennant and Conaghan [24].

234

235 Validity aspects to be tested

236

237 Structural validity testing appraises the fidelity of the scoring structure to the structure of the latent
238 construct domain. Using the first sample, the dimensionality of the questionnaire was measured to
239 ensure that the items were loading onto theoretically meaningful constructs. In line with the first
240 aim of this study (developing and testing a theory-driven reliable and valid questionnaire to assess
241 consultation-related reassurance in LBP) dimensionality Maps were run in Winsteps [23], which
242 assess how much variance is explained by the items as a whole, and provides estimates for clusters
243 which may represent separate dimensions. The Winsteps guide [25] recommends treating item
244 clusters with Eigenvalues of more than 2 as separate subscales, and subsequently running the
245 dimensionality maps again separately for the items which load more than 0.4 on the cluster, and for
246 the remaining items, and so on until no significant clusters remain. The results of each analysis were
247 investigated qualitatively (i.e. by checking the content of the items) to ensure that item clusters
248 were theoretically meaningful. Any sub-scales identified during this process were adhered to in
249 further analysis, described below.

250

251 Content validity refers to the relevance and representativeness of the items of the content upon
252 which they are based. Face validity for items had already been explored through expert review and
253 the use of patient advisory groups. We further tested the content validity of our measure according
254 to the Rasch model using item-measure correlations and standardised unweighted mean-squared fit
255 indices for each subscale separately. Item-measure correlations indicate how well scores on a
256 particular item are consistent with the average score across the remaining items. As advised by
257 Wolfe & Smith [18], correlations of 0.4 and above were considered satisfactory. Standardised
258 unweighted mean-squared fit indices evaluate individual items by comparing their observed and
259 expected values. This tells us how well each item 'fits' with the rest of the scale. An item with a
260 higher score suggests the presence of large residuals in the data, meaning that the item may not be
261 measuring the same construct as the rest of the items. Conversely, items with very low mean-
262 squared fit values indicate the data 'overfitting' the model, which could indicate redundancy in our
263 scale. Items with mean-squared fit values exceeding ± 2 were examined qualitatively to assess their
264 value to the scale, and removed as indicated, in line with the second aim of the study which was to
265 select a short version of the questionnaire by removing similar items to ensure our final instrument
266 is easily usable.

267

268 Differential Item Functioning (DIF) assesses whether items maintain their meaning across different
269 groups of respondents. In other words, whether individuals from different groups respond

270 differently to an item despite having the same ability level. DIF analyses were run across groups
271 according to education level (to ensure that the wording of the question did not discriminate
272 between those of higher and lower educational attainment) and physician gender (to assess
273 whether preconceived expectations of either gender's behaviour did not influence participants'
274 responses to the items). Items with DIF t-test scores of ± 2 or more were to be investigated
275 qualitatively.

276

277 Reliability was assessed in two ways, to further address the aim of the study in producing a valid and
278 reliable measure. First, the person- and item-separation and reliability indices built into the
279 Winsteps programme [23] were obtained within the Rasch model. Person separation is used to
280 classify people. Low person separation with a relevant person sample implies that the instrument
281 may not be sensitive enough to distinguish between high and low performers, and more items may
282 be needed. Item separation is used to verify the item hierarchy. Low item separation implies that the
283 person sample is not large enough to confirm the item difficulty hierarchy of the instrument.
284 Winsteps advises that a reliability coefficient of 0.5 is the minimum meaningful reliability, and 0.8 is
285 the minimum required for 'serious decision-making'. Therefore, subscales with a person- or item-
286 reliability score higher than 0.5 will be considered to show acceptable reliability, and subscales with
287 a person-or item-reliability score higher than 0.8 will be considered to show good reliability.

288

289 Secondly, correlational analysis comparing participants' scores at two time points (post-consultation
290 and one-week later) assessed the temporal reliability of the scale. The interval between responses is
291 important, because too short a gap can result in participants recalling and replicating their
292 responses, and too large a gap may result in recording real changes in patients' perceptions,
293 understanding and recall. We opted for a time interval of one week between receiving the responses
294 to the questionnaire and sending out the questionnaire again. An intraclass correlation coefficient
295 (ICC) is the most appropriate statistical method for continuous scores. Terwee et al [26] recommend
296 ICC agreement over ICC consistency because ICC agreement takes systematic error into account. This
297 requires at least 50 participants to provide two sets of responses to the scale [26]. This analysis was
298 conducted in SPSS version 21 [27], and coefficients of 0.7 or higher were considered acceptable [28].

299

300 External validity is the degree to which measures are related to external measures of the same,
301 similar, or other constructs. Spearman's Rho correlations were used to compare our scale with the

302 Consultation Satisfaction Questionnaire [CSQ, 11] and the Patient Enablement Instrument [PEI, 12,
303 13]. It was anticipated that the affective reassurance subscale would produce a positive correlation
304 of >0.4 with patient satisfaction, as measured by the CSQ. The cognitive reassurance subscale was
305 expected to produce a positive correlation of >0.4 with patient enablement, as measured by the PEI.
306 These predictions were derived from the theory upon which this questionnaire is based [2, 4], and
307 measuring these correlations further met the first aim of the study, to ensure that the questionnaire
308 was valid, reliable, and fit with current theory.

309

310 **Results**

311 **Participants**

312 One hundred and fifty-seven participants returned questionnaires for the first sample; 162 patients
313 provided data for sample 2. Patient characteristics are presented in Table 1.

314

315 **Structural Validity: Dimensionality Analyses**

316 A dimensionality map of the responses of Sample 1 on the entire scale revealed that it was not
317 unidimensional. See Figure 5 for a representation of the identified dimensions within the
318 questionnaire.

319 1. First, a major cluster was identified consisting of 9 items. A second dimensionality map of
320 this cluster showed that these items were also multidimensional, and separated them into
321 two clusters, one consisting of 3 data-gathering items and the other of 6 affective
322 reassurance items.

323 2. A dimensionality map of the remaining 21 items separated the other 4 data-gathering items
324 from the rest of the scale. *As depicted in Figure 5, the dimensionality analyses separated the*
325 *data-gathering items from the remainder of the item pool at the second stage. The three*
326 *items in the first cluster were:*

327 *4. Listen attentively while you were talking*

328 *5. Give you enough time to say everything you wanted to say*

329 *6. Ask questions to make sure he/she understood what you meant*

330 *The four items from the remaining pool were:*

331 *1. Ask about how your symptoms affect you in everyday life*

332 *2. Encourage you to voice your concerns regarding your symptoms*

333 3. *Ask you what you thought your symptoms might mean*

334 7. *Summarise what you had told them*

335 *As the key concepts underpinning data-gathering (demonstrating understanding of the*
336 *patient's problem; eliciting patients' concerns and finding out the whole story) were*
337 *represented across both of these clusters, they were assessed as not being qualitatively*
338 *different enough to warrant two subscales. Because the dimensionality analyses had*
339 *separated the data-gathering items from the items which concerned the later stages of the*
340 *consultation, the researchers made the decision to place all the items together in subsequent*
341 *analyses, with the understanding that analysis of fit indices would identify any items which*
342 *did not fit with the overall subscale.*

343 3. Next, dimensionality maps were run on the 23 data-giving items from the scale, and
344 provided three clusters. Out of 30 items, 24 mapped onto constructs hypothesised in the
345 model (highlighted in bold in Table 2). All of the items were retained at this stage for further
346 analysis. The items included in each newly identified subscale are presented in Table 2.

347

348 **Content Validity and Reliability**

349 Assessment using the principles of Rasch measurement was conducted on each subscale.

350 Data-Gathering

351 Seven items were entered into the Standardised unweighted mean-squared fit indices analysis and
352 calculation were carried omitting problematic items until both infit and outfit for the remaining
353 items fell within acceptable ranges. The final model, which included items 2, 4 and 7 (encourage you
354 to voice your concerns regarding your symptoms; listen attentively while you were talking; and
355 summarise what you had told them), showed good fit for all items and was used in the remainder of
356 analyses. Item-measure correlations were calculated for the reduced subscale, and were found to be
357 strong: 0.88, 0.80, and 0.88 for items 2, 4 and 7 respectively. This was then repeated in the second
358 sample, confirming the fit with all standardised unweighted mean-squared fit indices under the ± 2
359 threshold for problematic items, and item measure correlations ranging between 0.82- 0.92.

360 DIF statistics were calculated for items 2, 4 and 7 to assess whether different items were answered
361 differently by participants from different groups. For both samples, tests for education level and
362 physician level were non-significant.

363 Reliability was assessed for this subscale using Rasch person- and item-separation statistics and ICCs
364 comparing scores on the items one week after one another. For sample 1, the person separation was

365 2.08 (reliability coefficient 0.81), and the item separation was 8.67 (reliability coefficient 0.99),
366 indicating a good level of reliability. Reliability remained high for sample 2: person separation 2.26
367 (reliability coefficient 0.8); item separation 8.65 (reliability coefficient 0.99). The results for Average
368 Measures ICC with two-way mixed agreement are presented in Table 3. Correlations were all above
369 the acceptable level of 0.70, and so the subscale can be considered to have good test-retest
370 reliability.

371

372 Relationship building

373 Eight items were entered and the procedure described repeated. The final model, made up of items
374 7, 19 and 21 (show a genuine interest in your problem; put you at ease; and show that he/she
375 understood your concerns respectively), showed good fit for all items and was used in analysis of
376 sample 2. Item-measure correlations were calculated for the reduced subscale, and were found to
377 be 0.86, 0.91 and 0.91 for items 7, 19 and 21 respectively, suggesting that each of the items
378 correlated strongly with the final, reduced subscale. For sample 2, items 7 and 19 showed
379 standardised mean-squared fit indices outside of the acceptable ranges of ± 2 , suggesting the
380 presence of large residuals within the data. As removal of either of these items would leave only two
381 in the subscale, it was decided instead that all of the original Relationship-building items (see
382 previous page) would be re-entered using sample 2's data, to assess whether a different
383 combination of the items might better represent the construct. This model would then be re-
384 checked using the data from sample 1. The item-measure correlations for a subscale containing
385 items 4, 11, 15 and 6 were 0.87, 0.88, 0.82, and 0.90 respectively. When these items were entered
386 into Winsteps using sample 1's data, item 11 was misfitting (infit -2.3; outfit -2.4). This was removed,
387 and the remaining three items showed good fit for both samples. The three items in the second
388 reduced subscale (appear composed and level-headed; treat you politely; and show acceptance of
389 your concerns)Therefore, both subscales were analysed using the combined data from Sample 1 and
390 2 before a decision was reached on which to include in the final questionnaire. Both subscales
391 showed acceptable fit statistics and strong item-measure correlations.

392 DIF statistics showed that when separated by education level, or physician gender, variation was
393 evenly spread amongst groups for both subscales, with no significant t-test results.

394 For the first subscale, person- and item-reliability were both above the threshold for good reliability
395 (0.82 and 0.89, respectively). However, for the second subscale person reliability was 0.77, and

396 therefore failed to meet the standard for good reliability of >0.8, although item-separation was good
397 at 0.99. Test-retest reliability was strong for both subscales (see Table 3).

398

399 Overall, both potential subscales performed well when analysed using samples 1 and 2 combined.
400 However, the second subscale showed weaker person-separation than the first, which can be
401 indicative of a ceiling effect. As the items in the first subscale were felt to be more qualitatively
402 meaningful in the context of relationship-building, this subscale was included in the final
403 questionnaire.

404

405 Generic reassurance

406 Four Items were included in the Standardised unweighted mean-squared fit indices analysis of the
407 generic reassurance subscale. The final model, made up of items 9, 18 and 20 (tell you that you
408 should not be worried; tell you that everything would be fine; and reassure you that he/she had no
409 serious concerns about your back, respectively), showed good fit for all items and was used in
410 subsequent analyses. Item-measure correlations for the reduced subscale were 0.89, 0.90 and 0.85
411 for items 9, 18 and 20 respectively, suggesting that the items correlated well with overall subscale.
412 The subscale showed good fit when tested again with the data from sample 2. DIF statistics for both
413 samples sample 1 showed that variation was evenly spread amongst groups for education and
414 physician gender.

415 The generic reassurance subscale showed good reliability. For the first sample, person separation
416 was 2.12 (reliability coefficient 0.82) and the item separation was 4.15 (reliability coefficient 0.95).
417 For the second sample, the person separation was 2.07 (reliability coefficient 0.81) and the item
418 separation was 4.67 (reliability coefficient 0.96). ICC scores are shown in Table 5.15, and
419 demonstrate good test-retest reliability for this subscale (Table 3).

420 Cognitive reassurance

421 Eleven items were entered into the standardised unweighted mean-squared fit indices analysis. The
422 final model, made up of items 1, 12 and 23 (explain how the treatment offered would help with your
423 problem; make sure you understood what your treatment plan involves; and check you understood
424 the explanation he/she gave for your symptoms, respectively), showed good fit for all items and was
425 used in subsequent analyses. Item-measure correlations were 0.84, 0.81, and 0.84 for items 1, 12
426 and 23 respectively, suggesting that the items correlated well with the overall subscale. Fit statistics

427 and Item-measure correlations remained at acceptable levels using the data from sample 2. As for
428 the other sub-scales, education level and practitioner gender did not influence responses in either
429 sample.

430 Person- and item-separation indices were within acceptable ranges for sample 1: the person
431 separation was 2.04 (reliability coefficient 0.81) and the item separation was 2.48 (reliability
432 coefficient 0.86). For sample 2, the person separation was 1.82 (reliability coefficient 0.77) and the
433 item separation was 1.36 (reliability coefficient 0.65). Although the reliability scores for sample 2 fell
434 above the minimum meaningful level of 0.5, they failed to reach to acceptable standard of 0.8. ICCs,
435 however, were all strong for this subscale and indicate acceptable test-retest reliability (table X).

436

437 **External Validity**

438 All four subscales were significantly positively correlated with satisfaction and enablement, for both
439 samples (Table 4). The hypotheses that affective reassurance (in this case split into relationship-
440 building and generic reassurance) would show a positive correlation >0.4 with satisfaction, and that
441 cognitive reassurance would show a positive correlation >0.4 with enablement were both
442 supported. The final questionnaire is presented in table 5.

443

444

445 **Discussion**

446 The aims of this study were to develop and test a theory-driven reliable and valid questionnaire to
447 assess consultation-related reassurance in LBP. Data reduction, using Rasch analysis resulted in a 12
448 item questionnaire. Overall, the questionnaire performed well, with good content validity,
449 consistent responses across groups, and acceptable reliability. The final questionnaire represents
450 four distinct aspects of reassurance during consultations: data gathering, relationship building,
451 generic reassurance, and cognitive reassurance.

452 The four sub-categories map on to the model of reassurance proposed by Pincus et al (2013). The
453 first two, data gathering and relationship building can be considered to provide implicit reassurance,
454 while the latter can be conceptualised as explicit reassurance. According to Coia and Morley (1998),
455 relationship building and generic reassurance would fall into the category of affective reassurance,
456 combining verbal and non-verbal behaviours. Coia and Morley do not mention data gathering
457 behaviours, possibly because they consider these as attempts to elicit information about the
458 presenting problem, rather than attempts to understand the whole person's story, including their
459 concerns and the implications on their lives. As such, we consider that the items in the data-
460 gathering sub-scale also represent implicit reassurance, as they convey the patients perception that
461 they have had the opportunity to voice their concerns, and that they have been listened to.

462

463 **Strengths and limitations**

464 The split of the four subscales, whilst indeed different from the initial three-construct structure of
465 the overall item pool, we feel is a strength of the tool rather than a weakness. Two of the original
466 subscales were retained: data-gathering and cognitive reassurance; while the items which were at
467 first grouped together under the umbrella term 'affective reassurance', to represent all emotionally-
468 based attempts to reduce patients worry, were found to represent two distinct constructs:
469 relationship-building and generic reassurance. Within Coia and Morley's [4] conceptualisation of
470 reassurance, they describe affective reassurance as a combination of non-verbal cues which are
471 "largely synonymous with the doctor's manner" and direct verbal statements intended to
472 emotionally reassure. These two aspects of affective reassurance are represented within our final
473 questionnaire structure. Additionally, the separation of relationship-building behaviours from
474 generic reassurance statements maps to the distinction between implicit (unstated but perceived by
475 patients) and explicit (direct and often verbal) reassurance found in earlier qualitative work [5].

476 Therefore, the final, four-construct questionnaire provides more specificity in evaluating the model
477 than the original structure in which affective reassurance was considered a single construct.

478

479 As in all questionnaire development using data reduction techniques, we aim to produce a small set
480 of items that nonetheless captures the most salient items to describe the sub-scales in which they
481 are placed. For this reason our original pool of items includes replication and slightly different
482 voicing of the same item. We aim to exclude most of the items because we want to have a
483 questionnaire that is low burden to patients and therefore usable in research. One of the most
484 pressing problems in the study of psychosocial factors in pain (much like all research in patient
485 groups) is missing data and attrition due to inclusion of too many questionnaires, and questionnaires
486 that are unnecessarily long. The final 12 items included in this questionnaire all showed good fit with
487 the other items in their subscales as measured using standardised unweighted mean-squared indices
488 and item-measure correlations; acceptable reliability; no evidence of differential item functioning,
489 and good external validity when compared with established consultation outcome measures

490

491 Although the sub-scales were shown to have good reliability and validity, we have some concerns
492 about their ability to comprehensively capture all aspects of the consultation. For example,
493 relationship-building was one of the key skills extracted from the literature review, involving
494 emotion-based behaviours such as empathising, being supportive, and forming a bond. The benefits
495 of forming therapeutic relationships with patients are well-reported [e.g. 29, 30-33]. However, the
496 items produced by our analysis appears more superficial, reflecting the practitioners' ability to
497 convey confidence, act politely and acknowledge patients' concerns. Reliability was assessed for all
498 subscales using Rasch estimates of reliability and ICC scores comparing responses to the items given
499 one week apart. While test-retest reliability was demonstrated for all items and subscales, Rasch
500 estimations of reliability were mixed. Specifically, the cognitive reassurance subscale fell just short of
501 the higher standard of reliability (>0.8) when analysed using Sample 2's data. We acknowledge that
502 this is preliminary work, and that the questionnaire requires further validation to ensure full
503 confidence in its ability to reliably measure the different facets of reassurance.

504

505 The study utilised two separate samples for the analysis. While this enabled re-testing findings in a
506 new sample, it could be argued that both samples could be expected to perform similarly, as they
507 were drawn from the same population presenting to the same practices. However, the samples were
508 recruited from 43 general practices, in a large geographical spread and diverse socio-economical

509 catchment populations. This argument is supported by Differential Item Functioning (DIF) analysis,
510 which tests the different probability within groups of endorsing a particular item. All four subscales
511 showed no presence of DIF for either participant education level or physician gender, meaning that
512 responses did not differ significantly across respondents within different groups on these variables.
513 The absence of DIF for participant education is encouraging, as it is essential that a questionnaire is
514 understandable to people from all educational backgrounds [34]. Responses from participants
515 whose physicians had been of different genders were examined as there are documented
516 differences in the ways male and female physicians communicate with patients, with female
517 physicians more likely to engage in empathetic and partnership-building behaviours [35].
518 Additionally, physician gender has been shown to affect patient satisfaction outside of the effects of
519 patient characteristics and physician behaviours [36], suggesting that patients may hold expectations
520 for physicians of different genders which affect their perceptions of the care they receive. However,
521 all four subscales were resistant to these effects and remained consistent whether the physician in
522 question was male or female.

523

524 Scores on all four subscales were correlated with scores on established consultation measures for
525 satisfaction [CSQ, 11] and enablement [PEI, 12]. All showed significant positive correlations with
526 both instruments for both samples, demonstrating good external validity for the scale. Correlations
527 between the reassurance subscales and enablement were weaker than those between reassurance
528 and satisfaction. Reassurance represents a minimal intervention by GPs, and it may be the case that
529 more intensive intervention is required to enable some patients, particularly those who are
530 considered higher risk for chronicity [6]. Cognitive reassurance was related more strongly than the
531 other subscales to enablement. This finding supported both the hypothesis that the two would be
532 correlated, and the model of reassurance which posits that cognitive reassurance equips patients
533 with the knowledge and skills to manage their problem [2]. Surprisingly, although the generic
534 reassurance subscale was significantly correlated with satisfaction, it showed the weakest
535 correlations of the four subscales in both samples. It was predicted that this type of reassurance
536 would particularly increase satisfaction as it produces immediate reductions in anxiety [4]. The
537 relationship between generic reassurance and satisfaction remains problematic: contradictory
538 evidence was found in a systematic review of prospective cohorts in primary care (Pincus et al.,
539 2013), with three studies showing a positive association between the two, and two studies showing
540 negative associations.

541

542 An important limitation of the current study is the delay between consultation and recruitment, due
543 to electronic searches being carried out on a monthly basis. To truly capture participants'
544 perceptions of reassurance administration of the measure should take place at consultation exit. In
545 addition, participants were included in this sample with both acute and chronic low back pain. A
546 sample of acute cases only (i.e. people presenting with their first episode of LBP) would be more
547 informative, to avoid contamination from previous consultations.

548

549 Overall, the measure has demonstrated a good level of validity and generally acceptable reliability.
550 This is the first of its kind to focus specifically on reassurance for LBP in primary care settings, and
551 will enable researchers to further their understanding of what is reassuring within the context of low
552 back pain consultations, and how outcomes are affected by different types of reassurance.
553 Additionally, since reassurance is recommended by various guidelines for low back pain [e.g. 1, 37,
554 38] the measure may provide a useful training and audit tool for physicians. The new measure
555 requires testing in prospective cohorts, and would benefit from further validation against
556 ethnographic observation of consultations in real time.

557

558

559 **Declarations:**

560 List of Abbreviations:

561 LBP – Low Back Pain

562 PPI – Patient and Public Involvement

563 GP – General Practitioner

564

565 Ethics and consent to participate:

566 Ethical approval for this study was granted by the London City and East NHS Research Ethics

567 Committee. Potential participants were provided with a detailed information sheet about the study

568 which stated that consent would be implied by return of the questionnaire.

569

570 Consent to publish:

571 Not applicable.

572

573 Competing Interests:

574 The authors have no competing interests to declare

575

576 Funding:

577 NH was supported by a studentship from the Economic and Social Research Council's Doctoral

578 Training Centre.

579 This study was supported by a EUROSpine grant.

580 Neither funding body was involved in the design, in the collection, analysis, and interpretation of

581 data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

582

583 Authors' contributions:

584 NH and TP were both involved at all stages of questionnaire development, study design and
585 recruitment, and both helped to draft and finalise the manuscript.

586 NH conducted the statistical analyses.

587

588 Availability of data and materials:

589 All data supporting the findings presented here is contained within this manuscript.

590

591 Authors' information:

592 NH is a postgraduate research student, and this work forms part of her doctoral thesis.

593

594 Acknowledgements:

595 The authors would like to acknowledge and thank the expert researchers (S. Vogel, S.J. Taylor, M.

596 Underwood, and C. Bradley and team), PPI group and LBP patients who provided feedback on early

597 drafts of the questionnaire.

598

599

References

600

601 [1] Airaksinen O, Brox JJ, Cedraschi C, Hildebrandt J, Klüber-Moffett J, Kovacs F, Mannion AF,
602 Reis S, Staal JB, Ursin H *et al*: **Chapter 4. European guidelines for the management of**
603 **chronic nonspecific low back pain**. *Eur Spine J* 2006, **15 Suppl 2**:S192-300.

604 [2] Pincus T, Holt N, Vogel S, Underwood M, Savage R, Walsh DA, Taylor SJ: **Cognitive and**
605 **affective reassurance and patient outcomes in primary care: a systematic review**. *Pain*
606 2013, **154**(11):2407-2416.

607 [3] Shaw WS, Pransky G, Router DL, Winters T, Tveito TH, Larson SM: **The effects of patient-**
608 **provider communication on 3-month recovery from acute low back pain**. *J Am Board Fam*
609 *Med* 2011, **24**(1):16-25.

610 [4] Coia P, Morley S: **Medical reassurance and patients' responses**. *J Psychosom Res* 1998,
611 **45**(5):377-386.

612 [5] Holt N, Pincus T, Vogel S: **Reassurance during low back pain consultations with GPs: a**
613 **qualitative study**. *Br J Gen Pract* 2015, **65**(639):e692-701.

614 [6] Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, Konstantinou K, Main CJ,
615 Mason E, Somerville S *et al*: **Comparison of stratified primary care management for low**
616 **back pain with current best practice (STarT Back): a randomised controlled trial**. *Lancet*
617 2011, **378**(9802):1560-1571.

618 [7] Linton SJ, McCracken LM, Vlaeyen JW: **Reassurance: help or hinder in the treatment of**
619 **pain**. *Pain* 2008, **134**(1-2):5-8.

- 620 [8] Turk DC, Rudy TE, Sorkin BA: **Neglected topics in chronic pain treatment outcome studies: determination of success.** *Pain* 1993, **53**(1):3-16.
- 621
- 622 [9] Roland M, Morris R: **A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain.** *Spine (Phila Pa 1976)* 1983, **8**(2):141-144.
- 623
- 624
- 625 [10] Pincus T, Santos R, Breen A, Burton AK, Underwood M: **A review and proposal for a core set of factors for prospective cohorts in low back pain: a consensus statement.** *Arthritis Rheum* 2008, **59**(1):14-24.
- 626
- 627
- 628 [11] Baker R: **Development of a questionnaire to assess patients' satisfaction with consultations in general practice.** *Br J Gen Pract* 1990, **40**(341):487-490.
- 629
- 630 [12] Howie JG, Heaney DJ, Maxwell M: **Measuring quality in general practice (Royal College of General Practitioners Occasional Paper, 75).** London: Royal College of General Practitioners; 1997.
- 631
- 632
- 633 [13] Howie JG, Heaney DJ, Maxwell M, Walker JJ: **A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations.** *Fam Pract* 1998, **15**(2):165-171.
- 634
- 635
- 636 [14] Hays RD, Morales LS, Reise SP: **Item response theory and health outcomes measurement in the 21st century.** *Med Care* 2000, **38**(9 Suppl):II28-42.
- 637
- 638 [15] Reeve BB, Fayers P: **Applying item response theory modeling for evaluating questionnaire item and scale properties.** In: *Assessing quality of life in clinical trials: methods of practice 2.* 2 edn. Edited by Fayers P, Hays R. Oxford: Oxford University Press; 2005: 55-73.
- 639
- 640
- 641 [16] Reise SP, Ainsworth AT, Haviland MG: **Item response theory fundamentals, applications, and promise in psychological research.** *Current Directions in Psychological Science* 2005, **14**(2):95-101.
- 642
- 643
- 644 [17] DeVellis RF: **Classical test theory.** *Med Care* 2006, **44**(11 Suppl 3):S50-59.
- 645
- 646 [18] Wolfe EW, Smith EV, Jr.: **Instrument development tools and activities for measure validation using Rasch models: part II--validation activities.** *J Appl Meas* 2007, **8**(2):204-234.
- 647
- 648 [19] Wolfe EW, Smith EV, Jr.: **Instrument development tools and activities for measure validation using Rasch models: part I - instrument development tools.** *J Appl Meas* 2007, **8**(1):97-123.
- 649
- 650 [20] Bond TG, Fox CM: **Applying the Rasch model: Fundamental measurement in the human sciences,** 2 edn. New York: Routledge; 2007.
- 651
- 652 [21] Pampaka M, Williams J, Hutcheson G: **Measuring students' transition into university and its association with learning outcomes.** *British Educational Research Journal* 2012, **38**(6):1041-1071.
- 653
- 654
- 655 [22] Pampaka M, Williams J, Hutcheson G: **The association between mathematics pedagogy and learners' dispositions for university study.** *British Educational Research Journal* 2012, **38**(3):473-496.
- 656
- 657
- 658 [23] Linacre JM: **Winsteps® Rasch measurement computer program.** In., 3.81.0 edn. Beaverton, Oregon: Winsteps.com; 2014.
- 659
- 660 [24] Tennant A, Conaghan PG: **The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?** *Arthritis Rheum* 2007, **57**(8):1358-1362.
- 661
- 662
- 663 [25] **Dimensionality investigation - an example**
- 664 [<http://www.winsteps.com/winman/multidimensionality.htm>] Accessed on 15 June 2015.
- 665 [26] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC: **Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist.** *Qual Life Res* 2012, **21**(4):651-657.
- 666
- 667
- 668 [27] IBM Corp: **IBM SPSS Statistics for Windows, Version 21.0.** In., 21.0 edn. Armonk, NY: IBM; 2012.
- 669

- 670 [28] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC:
671 **Quality criteria were proposed for measurement properties of health status**
672 **questionnaires.** *J Clin Epidemiol* 2007, **60**(1):34-42.
- 673 [29] Hall AM, Ferreira PH, Maher CG, Latimer J, Ferreira ML: **The influence of the therapist-**
674 **patient relationship on treatment outcome in physical rehabilitation: a systematic review.**
675 *Phys Ther* 2010, **90**(8):1099-1110.
- 676 [30] Hall JA, Roter DL, Katz NR: **Meta-analysis of correlates of provider behavior in medical**
677 **encounters.** *Med Care* 1988, **26**(7):657-675.
- 678 [31] Hsiao CJ, Boulton C: **Effects of quality on outcomes in primary care: a review of the literature.**
679 *Am J Med Qual* 2008, **23**(4):302-310.
- 680 [32] Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, Till J: **Doctor-patient**
681 **communication: the Toronto consensus statement.** *BMJ* 1991, **303**(6814):1385-1387.
- 682 [33] Makoul G: **Essential elements of communication in medical encounters: the Kalamazoo**
683 **consensus statement.** *Acad Med* 2001, **76**(4):390-393.
- 684 [34] Mandal A, Eaden J, Mayberry MK, Mayberry JF: **Questionnaire surveys in medical research.**
685 *J Eval Clin Pract* 2000, **6**(4):395-403.
- 686 [35] Jefferson L, Bloor K, Birks Y, Hewitt C, Bland M: **Effect of physicians' gender on**
687 **communication and consultation length: a systematic review and meta-analysis.** *J Health*
688 *Serv Res Policy* 2013, **18**(4):242-248.
- 689 [36] Bertakis KD, Franks P, Azari R: **Effects of physician gender on patient satisfaction.** *J Am Med*
690 *Wom Assoc* 2003, **58**(2):69-75.
- 691 [37] Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C: **An updated overview of**
692 **clinical guidelines for the management of non-specific low back pain in primary care.** *Eur*
693 *Spine J* 2010, **19**(12):2075-2094.
- 694 [38] van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, Koes B, Laerum E,
695 Malmivaara A: **Chapter 3. European guidelines for the management of acute nonspecific**
696 **low back pain in primary care.** *Eur Spine J* 2006, **15** Suppl 2:S169-191.
- 697
- 698
- 699

700 Figure 1: Conceptual map of data gathering

701

702 Figure 2: Conceptual map of Cognitive Reassurance

703

704 Figure 3: Conceptual map of Affective Reassurance

705

706 Figure 4: Collection and analysis of data

707

708 Figure 5: Dimensionality Mapping results

709

710 Table 1: Participant Characteristics

	Sample 1	Sample 2
Average Age	56.63 (SD 16.64)	53.52 (SD 16.08)
Gender	63.9% female 36.1% male	63.4% female 36.6% male
Length of current episode	33.8% <1 month 23.0% 1-3 months 11.5% 4-6 months 14.2% 7 months – 3 years 17.6% >3 years	24.1% <1 month 27.2% 1-3 months 11.4% 4-6 months 23.4% 7 months – 3 years 13.9% >3 years
Number of consultations for this episode	47.9% none 31.9% 1-2 14.3% 3-10 5.9% >10	54.4% none 30.9% 1-2 12.5% 3-10 2.2% >10
Work status	53.9% employed (full or part time) 35.7% retired 3.9% looking after home/family 1.9% unemployed (health reasons) 2.6% unemployed (other) 1.9% student	56.2% employed (full or part time) 32.1% retired 3.1% looking after home/family 3.7% unemployed (health reasons) 1.9% unemployed (other) 3.1% student
Education level	49.0% obtained higher education degree/certification 18.1% obtained A levels or equivalent 32.9% left school at or before 16	44.0% obtained higher education degree/certification 20.7% obtained A levels or equivalent 35.3% left school at or before 16
Marital status	65.8% married/civil partnership 7.7% cohabiting 7.7% single 9.7% divorced 6.5% widowed 2.6% other	57.8% married/civil partnership 9.9% cohabiting 14.9% single 12.4% divorced 5.0% widowed
Physician type	99.3% GP	96.3% GP

	0.7% nurse practitioner	3.8% nurse practitioner
Physician gender	52.9% male 47.1% female	50.9% male 49.1% female
First episode?	26.1% yes 73.9% no	27.2% yes 72.8% no
Average pain intensity in the last week (/10)	7.14 (SD 2.02)	7.06 (SD 2.06)
RMDQ score (/24)	10.34 (SD 5.73)	10.10 (SD 5.98)

711

712

Cluster 1 (Data-Gathering)	Cluster 2 (Relationship-Building)	Cluster 3 (Generic Reassurance)	Cluster 4 (Cognitive Reassurance)
1. Ask about how your symptoms affect you in your everyday life 4.10 (1.92)	4. Appear composed and level-headed 6.06 (1.04)	9. Tell you that you should not be worried 3.96 (2.05)	1. Explain how the treatment offered would help with your problem 4.51 (1.78)
2. Encourage you to voice your concerns regarding your symptoms 4.50 (1.82)	11. Seem friendly and approachable 5.82 (1.31)	16. Give a clear timescale for when your symptoms should improve 3.88 (2.15)	2. Give you a clear explanation for your symptoms 4.36 (1.88)
3. Ask you what you thought your symptoms might mean 3.54 (1.97)	7. Show a genuine interest in your problem 5.38 (1.61)	18. Tell you that everything would be fine 3.52 (2.09)	3. Chat with you informally 4.89 (4.47)
4. Listen attentively while you were talking 5.75 (1.27)	15. Treat you politely 6.24 (1.01)	20. Reassure you that he/she had no serious concerns about your back 4.38 (2.02)	5. Encourage you to be optimistic 4.75 (1.71)
5. Give you enough time to say everything you wanted to say 5.56 (1.50)	6. Show acceptance of your concerns 5.30 (1.56)		8. Give you a choice of treatment options 3.72 (2.12)
6. Ask questions to make sure he/she understood what you meant 5.18 (1.72)	19. Put you at ease 5.13 (1.79)		10. Seem pleased with how you had managed your symptoms so far 4.26 (1.89)
7. Summarise what you had told them 4.77 (1.86)	13. Check that you agreed with the treatment plan 4.85 (1.97)		12. Make sure you understood what your treatment plan involves 4.95 (1.94)
	21. Show that he/she understood your concerns 5.12 (1.80)		14. Assure you that you could control your problem 4.22 (2.01)
			17. Explain your symptoms in relation to your concerns 4.40 (2.04)
			22. Consider your lifestyle and needs in planning your treatment 4.18 (2.13)
			23. Check you understood the explanation he/she gave for your symptoms 4.65 (1.96)

714 Items highlighted in bold are those which mapped directly to the theoretical constructs in the model.

715 Numbers given in italics: *mean (SD)*

716

717

718 Table 3: Intraclass Correlation Coefficients (ICCs) for all subscales

	ICC Sample 1	ICC Sample 2
<i>Data gathering</i>		
Item 2	0.85, n=75(74,74)	0.82, n=68(67,67)
Item 4	0.83, n=74(73,73)	0.70, n=67(66,66)
Item 7	0.77, n=74(73,73)	0.75, n=68(67,67)
Whole subscale	0.90, n=76(75,75)	0.81, n=68(67,67)
<i>Relationship building (Subscale 1)</i>		
Item 7	0.87, n=155(154,154)	
Item 19	0.84, n=155(154,154)	
Item 21	0.88, n=154(153,153)	
Whole subscale	0.93, n=153(152,152)	
<i>Relationship-building (Subscale 2)</i>		
Item 4	0.78, n=156(155,155)	
Item 6	0.80, n=156(155,155)	
Item 15	0.86, n=156(155,155)	
Whole subscale	0.88, n=156(155,155)	
<i>Generic reassurance</i>		
Item 9	0.87, n=71(70,70)	0.82, n=68(67,67)
Item 18	0.90, n=68(67,67)	0.83, n=66(65,65)
Item 20	0.89, n=73(72,72)	0.77, n=68(67,67)
Whole subscale	0.91, n=73(72,72)	0.87, n=68(67,67)
<i>Cognitive reassurance</i>		
Item 1	0.82, n=72(71,71)	0.82, n=65(64,64)
Item 12	0.82, n=71(70,70)	0.79, n=65(64,64)
Item 23	0.85, n=72(71,71)	0.79, n=66(65,65)
Whole subscale	0.82, n=73(72,72)	0.88, n=66(65,65)

719

720

721 Table 4: Correlations between Reassurance Subscales and Satisfaction and Enablement Scales

	Total Satisfaction Score (CSQ)	Total enablement score (PEI)
<i>Sample 1</i>		
Data Gathering, n=156	0.71*	0.43*
Generic Reassurance, n=151	0.54*	0.42*
Cognitive Reassurance, n=156	0.80*	0.48*
<i>Sample 2</i>		
Data Gathering, n=162	0.77*	0.43*
Generic Reassurance, n=160	0.45*	0.46*
Cognitive Reassurance, n=162	0.76*	0.52*
<i>Combined Samples</i>		
Relationship-building Subscale 1, n=312	0.81*	0.52*
* correlation significant at p<0.05		

722

723

724 Table 5: Final reassurance questionnaire

Data-gathering subscale	Relationship-building subscale	Generic reassurance subscale	Cognitive reassurance subscale
<i>To what extent did the physician ...</i>			
Encourage you to voice your concerns regarding your symptoms	Show a genuine interest in your problem	Tell you that you should not be worried	Explain how the treatment offered would help with your problem
Listen attentively while you were talking	Put you at ease	Tell you that everything would be fine	Make sure you understood what your treatment plan involves
Summarise what you had told them	Show that he/she understood your concerns	Reassure you that he/she had no serious concerns about your back	Check you understood the explanation he/she gave for your symptoms