

1 **Development and validation of the Symptom Burden Questionnaire™ for Long**  
2 **Covid: A Rasch analysis**

3  
4  
5  
6  
7  
8  
9

Sarah E Hughes<sup>1,2,3,4\*</sup>, Shamil Haroon<sup>4</sup>, Anuradhaa Subramanian<sup>4</sup>, Christel McMullan<sup>1,5,7</sup>, Olalekan L Aiyegbusi<sup>1,2,3,4,6</sup>, Grace M Turner<sup>1,4</sup>, Louise Jackson<sup>4</sup>, Elin Haf Davies<sup>8</sup>, Chris Frost<sup>8</sup>, Gary McNamara<sup>8</sup>, Gary Price<sup>1</sup>, Karen Matthews<sup>9,10</sup>, Jennifer Camaradou<sup>10,11</sup>, Jane Ormerod<sup>12</sup>, Anita Walker<sup>1</sup>, Melanie J Calvert<sup>1-6,13-15</sup>

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

<sup>1</sup>Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, UK

<sup>2</sup>National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) – West Midlands, Birmingham, UK

<sup>3</sup>Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, UK

<sup>4</sup>Institute of Applied Health Research, University of Birmingham, Birmingham, UK

<sup>5</sup>NIHR Surgical Reconstruction and Microbiology Research Centre, University Hospital Birmingham and University of Birmingham, UK

<sup>6</sup>NIHR Birmingham Biomedical Research Centre, University Hospital Birmingham and University of Birmingham, Birmingham, UK

<sup>7</sup>Centre for Trauma Science Research, University of Birmingham, UK

<sup>8</sup>Aparito Limited

<sup>9</sup>LongCOVIDSOS, UK

<sup>10</sup>Patient and Public Involvement, Therapies for Long COVID (TLC) study, University of Birmingham, UK

<sup>11</sup>Citizen Partner, COVID-END (COVID-19 Evidence synthesis Network to support Decision-making (WHO partner)

<sup>12</sup>Long COVID Scotland, UK

<sup>13</sup>DEMAND Hub, University of Birmingham, Birmingham, UK

<sup>14</sup>Health Data Research UK, London, UK

<sup>15</sup>UK SPINE, University of Birmingham, Birmingham, UK

\*Corresponding author:

E-mail: [s.e.hughes@bham.ac.uk](mailto:s.e.hughes@bham.ac.uk)

39 **ABSTRACT**

40 **Objective**

41 To describe the development and initial validation of a novel patient-reported outcome measure of  
42 Long COVID symptom burden, the Symptom-Burden Questionnaire for Long COVID (SBQ™-LC).

43 **Method and Findings**

44 This multi-phase, prospective mixed-methods study took place between April and August 2021 in the  
45 United Kingdom (UK). A conceptual framework and initial item pool were developed from published  
46 systematic reviews. Further concept elicitation and content validation was undertaken with adults with  
47 lived experience (n = 13) and clinicians (n = 10), and face validity was confirmed by the Therapies for  
48 Long COVID Study Patient and Public Involvement group (n = 25). The draft SBQ™-LC was field  
49 tested by adults with self-reported Long COVID recruited via social media and international Long  
50 COVID support groups (n = 274). Thematic analysis of interview and survey transcripts established  
51 content validity and informed construction of the draft questionnaire. Rasch analysis of field test data  
52 guided item and scale refinement and provided evidence of the final SBQ™-LC's measurement  
53 properties. The Rasch-derived SBQ™-LC is composed of 17 independent scales with promising  
54 psychometric properties. Respondents rate symptom burden during the past 7-days using a  
55 dichotomous response or 4-point rating scale. Each scale provides coverage of a different symptom  
56 domain and returns a summed raw score that may be converted to a linear (0 – 100) score. Higher  
57 scores represent higher symptom burden.

58 **Conclusions**

59 The SBQ™-LC is a comprehensive patient-reported assessment of Long COVID symptom burden  
60 developed using modern psychometric methods. It measures symptoms of Long COVID important to  
61 individuals with lived experience and may be used to evaluate the impact of interventions and inform  
62 best practice in clinical management.

## 63 INTRODUCTION

64 Since its emergence in 2019, the global COVID-19 pandemic has caused more than 250 million  
65 infections and over five million deaths [1]. For many individuals, the infection is mild and short-lived;  
66 however, a proportion of those infected continue to experience, or go on to develop, long-term  
67 symptoms that persist beyond the acute phase of infection. These persistent symptoms are known  
68 collectively as “post-acute sequelae of COVID-19 (PACS)”, “post-acute COVID-19”, “post-COVID-19  
69 syndrome”, or “Long COVID” [2].

70

71 Symptom burden may be defined as the “subjective, quantifiable prevalence, frequency, and severity  
72 of symptoms placing a physiologic burden on patients and producing multiple negative, physical, and  
73 emotional patient responses” [3]. The symptoms reported by individuals with Long COVID are  
74 heterogenous, affecting multiple organ systems with fatigue, dyspnea, and impaired concentration  
75 being among the most prevalent symptoms [4–6]. Symptoms may be persistent, cyclical, or episodic  
76 in their presentation and can pose a significant burden for affected individuals with negative  
77 consequences for work capability, functioning, and quality of life [7–9]. There is a growing body of  
78 research on the prevalence, incidence, co-occurrence, and persistence of Long COVID signs and  
79 symptoms [4,5,7,10–13]. To date, these data have largely been collected using bespoke, cross-  
80 sectional survey tools due, in part, to the limited availability of condition-specific, validated self-report  
81 instruments.

82

83 Patient-reported outcomes (PROs) are measures of health reported directly by the patient without  
84 amendment or interpretation by a clinician or anyone else [14]. Validated PRO instruments developed  
85 specifically for Long COVID that address the complex, multi-factorial nature of the condition are  
86 needed urgently to further understanding of Long COVID symptoms and underlying pathophysiology,  
87 support best practice in the clinical management of patients, and to evaluate the safety, effectiveness,  
88 acceptability, and tolerability of interventions [15–17]. Validated PRO instruments have recently been  
89 developed to measure the global impact of Long COVID and there are several unvalidated screening  
90 tools, surveys, and questionnaires also available [18,19]. However, individuals living with Long COVID  
91 have suggested that existing self-report measures fail to capture the breadth of experienced  
92 symptoms [9,20,21]. To address the need for a comprehensive measure of self-reported symptom

93 burden specific to Long COVID, this study used Rasch analysis to develop and validate, in  
94 accordance with US Food and Drug Administration (FDA) guidance, a novel PRO instrument, the  
95 Symptom Burden Questionnaire™ for Long COVID (SBQ™-LC) [14,22].

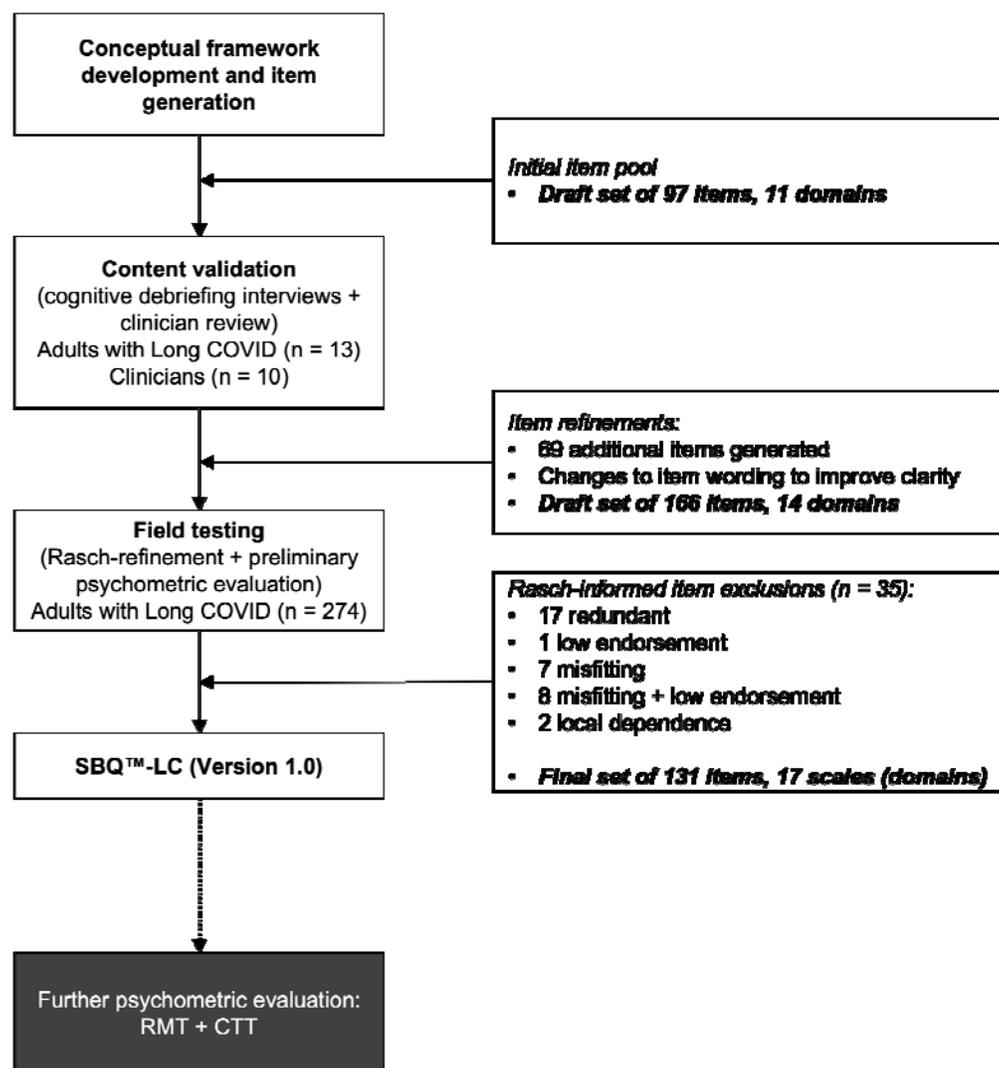
96

## 97 **METHODS**

### 98 **Setting and study design**

99 This multi-phase, prospective mixed-methods study (Fig 1) was nested within the Therapies for Long  
100 COVID for non-hospitalized individuals (TLC) study (ISRCTN Reference: 15674970) and took place  
101 from 14 April to 1 August 2021 [23]. A favorable ethical opinion was granted by the University of  
102 Birmingham Research Ethics Committee (Reference: ERN\_21-0191).

103



104

105 **Figure 1.** Study flow diagram showing the development of SBQ™-LC (Version 1.0)

## 106 **Study population**

107 Content validation was undertaken with adults with Long COVID recruited from the TLC study's  
108 Patient and Public Involvement and Engagement (PPIE) group and clinicians recruited from the TLC  
109 study team and Long COVID research studies based in the United Kingdom. The field test population  
110 included adults with self-reported Long COVID. Participants were aged 18 years or older who could  
111 self-complete the SBQ™-LC in English. No exclusion criteria relating to duration of Long COVID  
112 symptoms, hospitalization for SARS CoV-2, or vaccination status were applied. A minimum sample  
113 size of 250 respondents was prespecified for field testing. In Rasch analysis, a sample of 250  
114 respondents provides 99% confidence that item calibrations and person measures are stable within ±  
115 0.50 logits [24].

116

## 117 **Symptom coverage and existing PROs**

118 The SBQ™-LC's conceptual framework was developed from systematic literature reviews of Long  
119 COVID symptoms [4,25]. Existing symptom measures (n = 6) with good face validity in the context of  
120 Long COVID were reviewed to establish whether a new PRO instrument of symptom burden was  
121 needed [26–31]. When mapped to the conceptual framework, symptom coverage of these measures  
122 ranged from 25.0% to 64.1% (mean = 42.4% SD 16.8). This finding suggested complete coverage of  
123 Long COVID symptoms could not be guaranteed using existing measures, providing justification for  
124 the development of the SBQ™-LC.

125

## 126 **Study Procedures**

### 127 Content validation

128 Content validation involved cognitive debriefing with adults with Long COVID and an online clinician  
129 survey. Cognitive debriefing interviews to ascertain the relevance, comprehensiveness,  
130 comprehensibility, and acceptability of the SBQ™-LC's items for adults with lived experience of Long  
131 COVID were conducted via videoconferencing and recorded. The clinician survey was administered  
132 using the survey software application, SmartSurvey [32]. Verbatim transcripts of the interview  
133 recordings, field notes, and free text comments from the clinician survey were analyzed qualitatively  
134 using thematic analysis and a pre-specified framework to identify problems with the relevance,  
135 comprehensiveness, clarity, and acceptability of the items [33,34]. Participants with lived experience

136 of Long COVID were asked to identify additional symptoms not captured by the initial item pool.

137 Findings informed revisions which were tracked per item using an Excel spreadsheet. A draft version

138 of the SBQ™-LC was constructed and sent forward for field testing.

#### 139 Field testing

140 The Aparito Atom5™ platform is a regulated (ISO13485, ISO/IEC 27001:2013 Accreditation, FDA

141 CFR21 Part 11 compliant) software platform that provides remote data collection and real-time patient

142 monitoring via a smartphone application and integration with wearable devices [35]. The draft SBQ™-

143 LC and a demographic questionnaire were programmed onto Atom5™ for delivery as an electronic

144 patient-reported outcome (ePRO). Participants were recruited via social media advertisements posted

145 on Twitter and Facebook by the study team and through other support group platforms and website

146 registrations via collaborating Long COVID support groups based in the UK. Interested individuals

147 clicked on a URL connecting them to a study-specific website where they could read detailed

148 information about the study, provide their informed consent, and download the Atom5™ app to their

149 mobile device (i.e., smartphone or tablet). Participants accessed the SBQ™-LC and demographic

150 questionnaire via a unique QR code and once the completed questionnaires were submitted,

151 participants could delete the app from their phone. The anonymised field test data were downloaded

152 securely from Atom5™ for analysis.

153

#### 154 **Statistical Analyses**

155 STATA (Version 16) was used to clean and prepare the data and SPSS (Version 28.0) was used for

156 descriptive data analyses. Rasch analysis was conducted on the field test data to refine the SBQ™-

157 LC and assess its psychometric properties. A Rasch analysis is the formal evaluation of a PRO

158 instrument against the Rasch measurement model. The Rasch model is a mathematical ideal that

159 specifies a set of criteria for the construction of interval-level measures from ordinal data [36]. It is a

160 probabilistic model that specifies that a person's response to an item is only governed by the person

161 and the location of the item on a shared scale measuring the latent trait. The probability that a person

162 will endorse an item is a logistic function of the difference between a person's trait level (expressed as

163 person ability) and the amount of trait expressed by the item (expressed as item difficulty) [37]. Rasch

164 analysis enabled the SBQ™-LC to be constructed as a modular PRO instrument (i.e., a multi-domain

165 item bank) with linear, interval-level measurement properties. These properties render Rasch-

166 developed PROs suitable for use with individual patients as well as for group-level comparisons,  
167 permit direct comparisons of scores across domains, and facilitate the construction of alternative test  
168 formats (i.e., short forms and computer adaptive tests (CATs) [38,39].  
169  
170 Rasch analyses were carried out using Winsteps software (Version 5.0.5) and the Partial Credit  
171 Model (PCM) for polytomous data. The PCM was selected as the question wording and rating scale  
172 categories varied across items. Joint Maximum Likelihood Estimation (JMLE) in Winsteps enabled  
173 parameter estimation where data were missing [40]. Items scored as “not applicable” were considered  
174 ‘missing’ and did not contribute to item estimates. Misfitting response patterns (e.g., arising from  
175 respondent guessing or other unexpected behaviour) have been shown to result in biased item  
176 estimates with detrimental impacts for model fit [41]. Therefore, as is customary in Rasch analysis,  
177 person fit statistics were appraised and persons with misfitting response patterns (i.e., outfit MnSq  
178 logits > 2.0 logits) were removed iteratively and item parameters re-estimated until evidence of item  
179 parameter stability was observed [37,42].  
180  
181 Rating scale functioning for individual items was assessed against the following criteria: 1) items  
182 oriented in the same direction as a check for data entry errors (i.e., appraisal of point-measure  
183 correlations), 2) average category measures advance (i.e., higher categories reflect higher measures;  
184 3) category outfit mean-square (MnSq) values do not exceed 2.0 logits (i.e., as an indicator of  
185 unexpected randomness in the model); and 4) each category endorsed by a minimum of ten  
186 respondents [43]. If an item’s rating scale failed to meet these criteria, adjacent categories were  
187 combined or the item was removed. Category probability curves (CPCs) provided a graphical  
188 representation as further evidence of rating scale functioning.  
189  
190 To confirm model fit, Rasch analyses (including appraisal of unidimensionality, local independence,  
191 and individual item fit statistics) were completed iteratively as items were removed or grouped to  
192 create new scales. Person reliability and separation indices and scale-to-sample targeting were also  
193 evaluated. Targeting examines the correspondence between items and persons and, for a well-  
194 targeted scale, the items in a scale should be spaced evenly across a reasonable range of the scale  
195 and correspond to the range of the construct experienced by the sample. Person reliability examines

196 the reproducibility of relative measure location and person separation provides a measure of the  
197 number of distinct levels of person ability (i.e., symptom burden) that can be distinguished by a scale  
198 [44]. Cronbach's alpha, as a measure of the internal consistency reliability, was computed for each  
199 scale [45]. The list of parameters evaluated in the development and validation of the SBQ™-LC, along  
200 with acceptability criteria, are described in Table 1.

**Table 1.** Rasch measurement properties and acceptability criteria

| Measurement property  | Definition/Aim of evaluation  | Acceptability criteria for fit  |
|---|---|---|
| <b>Valid measurement model</b>  | To identify a set of items which effectively measure the target construct of symptom burden in Long COVID (i.e., fulfil the axioms of fundamental measurement permitting the construction of interval level scales).  | <ul style="list-style-type: none"> <li>• Unidimensionality: PCAR<sup>a</sup> highest eigenvalue of first residual contrast &lt; 2.0; disattenuated correlations &gt; 0.70</li> <li>• Local item independence: Residual correlation &lt; 0.40</li> <li>• Fit statistics: Outfit/infit MnSq values within 0.5 – 1.5 logits</li> <li>• Point measure correlations &gt; 0.40</li> <li>• Rating scales: scale oriented with latent variable, categories advance monotonically, category fit statistics: Outfit MnSq values &lt; 2.0 logits, uniform category endorsement.</li> </ul> |
| <b>Reliability</b><br>Rasch reliability<br><br><br><br><br><br><br><br><br><br>Internal consistency reliability | <p>Rasch-based reliability is the share of true variance of the total observed variance of the measure.</p> <p>Person separation index (PSI) is the number of distinct levels of person ability (i.e., symptom burden) that can be identified by the measure.</p> <p>The extent to which items comprising a scale measure the same construct (i.e., the degree of homogeneity or relatedness among the items of a scale).</p> | <ul style="list-style-type: none"> <li>• Rasch-based reliability:<br/> <math>r \geq 0.70</math> acceptable<br/> <math>r \geq 0.80</math> good<br/> <math>r \geq 0.90</math> excellent</li> <li>• PSI<sup>b</sup> = 1.5 – 2.0 acceptable level of separation<br/>           PSI = 2.0 – 3.0 good level of separation<br/>           PSI <math>\geq 3.0</math> excellent level of separation</li> <li>• Cronbach's alpha range <math>\geq 0.70</math></li> </ul>  |
| <b>Targeting</b>  | The extent to which the range of the variable measured by the scale matches the range of that variable in the study sample.   | <ul style="list-style-type: none"> <li>• Item-person map</li> <li>• Acceptable targeting is demonstrated by close correspondence of the person mean with the item mean for a scale (i.e., <math>\pm 1.0</math> logits from the mean of zero).</li> </ul>  |

<sup>a</sup>Principal component analysis of the residuals; <sup>b</sup>PSI = person separation index

## 1 **RESULTS**

### 2 **Item development and content validation**

3 An initial item pool (n = 97) was constructed, guided by the conceptual framework developed from the  
4 published literature. Clinician review of the item pool informed changes to improve item clarity.  
5 Content validity was confirmed by individuals with lived experience of Long COVID (n = 13) in two  
6 rounds of cognitive debriefing interviews. Participants were of white ethnicity, ten (76.9%) were  
7 female, and ranged in age from 20 – 60 years. Cognitive debriefing identified gaps in symptom  
8 coverage, resulting in the generation of 69 new items. Thematic analysis classified problems with  
9 draft items' relevance, comprehensiveness, clarity, acceptability. S1 Table presents key themes,  
10 together with exemplar quotations, from the thematic analysis.

11  
12 The draft SBQ™-LC included 166 items (155 symptoms + 11 interference items) and an *a priori*  
13 theoretical domain classification comprised of 14 domains, each constructed as an independent  
14 scale. Items utilized a 7-day recall period and burden was measured using a dichotomous “yes/no”  
15 response or a 5-point rating scale measuring severity, frequency, or interference. Higher scores  
16 represented greater symptom burden. Face validity of the draft SBQ™-LC was confirmed by TLC  
17 study's PPI group.

18

### 19 **Readability**

20 Readability, measured using Flesch-Kincaid Reading Grade Level Test and the SMOG Index score,  
21 was calculated using the web-based application Readable [46–48]. The American Medical  
22 Association (AMA) and the National Institutes of Health (NIH) recommend readability of patient  
23 materials should not exceed a sixth-grade reading level (a Flesch-Kincaid Reading Grade Level of  
24 score of 6.0) [49]. The SBQ™-LC's Flesch-Kincaid Grade Level score was 5.33 and SMOG index  
25 score was 8.27. Text with a SMOG index score ranging between 7 and 9 would be understood by  
26 93% of adults in the UK [50].

27

### 28 **Field testing**

#### 29 Participants

30 Over a two-week period, 906 questionnaires were delivered and 330 responses received. Fifty-six  
31 submissions were incomplete and excluded from the analyses. The final sample included 274  
32 complete responses, a response rate of 30.2%. Respondents' ages ranged from 21 to 70 years, with  
33 a mean (SD) age of 45.1 (10.1) years. The sample included 240 (87.6%) females and 253 (92.3%)  
34 respondents were white. Two hundred and eleven (77.0%) participants reported three or more  
35 comorbidities. One hundred and twenty-nine (47.1%) participants reported a diagnosis of SARS-CoV-  
36 2 that was confirmed by a positive PCR test and 22 (8.0%) respondents reported confirmation of  
37 SARS-CoV-2 by a positive lateral flow test. Eleven participants (4.0%) had been hospitalized due to  
38 SARS-CoV-2 and 70 (25.5%) had received one dose of vaccine while 187 (68.2%) had received two  
39 doses. In total, 153 (55.8%) respondents were in either full-time or part-time employment (Table 2).

40 **Table 2.** Demographic characteristics of the field test sample (n = 274)  
41

|  | n (%)            |
|--|------------------|
| <b>Age (n = 263 reported)</b>  |                  |
| Mean   | 45.1 years       |
| SD   | 10.1 years       |
| Range  | 21 – 70 years    |
| <b>Gender</b>  |                  |
| Female   | 240 (87.6)       |
| Male   | 34 (12.4)        |
| <b>Ethnicity</b>   |                  |
| Asian or Asian British   | 7 (2.6)          |
| Black, African, Caribbean, or Black British                                | 3 (1.1)          |
| Mixed or Multiple ethnic groups  | 11 (4.0)         |
| Other ethnic group   | 0 (0.0)          |
| White  | 253 (92.3)       |
| <b>Occupational status</b>   |                  |
| Caregiver  | 3 (1.1)          |
| Employed full-time   | 114 (41.6)       |
| Employed but currently not working   | 51 (18.6)        |
| Employed part-time   | 39 (14.2)        |
| Furloughed   | 7 (2.6)          |
| In full-time education   | 3 (1.1)          |
| Retired  | 6 (2.2)          |
| Voluntary work   | 3 (1.1)          |
| Unemployed   | 20 (7.3)         |
| Other  | 28 (10.2)        |
| <b>Positive PCR Test for SARS-Co-V-2 infection</b>                         |                  |
| Yes  | 129 (47.1)       |
| No   | 137 (50.0)       |
| Don't know   | 8 (2.9)          |
| <b>Positive lateral flow test for SARS-Co-V-2 infection</b>                |                  |
| Yes  | 22 (8.0)         |
| No   | 227 (82.8)       |
| Don't know   | 25 (9.1)         |
| <b>Hospitalized for SAR-Co-V-2 infection</b>                               |                  |
| Yes  | 11 (4.0)         |
| No   | 263 (96.0)       |
| <b>Admitted to intensive care unit due to SARS-Co-V-2 infection</b>        |                  |
| Yes  | 0 (0.0)          |
| No   | 274 (100.0)      |
| <b>Attended hospital emergency department due to SARS-Co-V-2 infection</b> |                  |
| Yes  | 111 (40.5)       |
| No   | 163 (59.5)       |
| <b>Vaccine status as of (XX-Jun-2021)</b>                                  |                  |
| Two doses  | 187 (68.2)       |
| One dose   | 70 (25.5)        |
| No dose  | 17 (6.2)         |
| <b>Received shielding letter from UK government<sup>a</sup></b>            |                  |
| Yes  | 12 (4.4)         |
| No   | 262 (95.6)       |
| <b>Care home resident</b>  |                  |
| Yes  | 3 (1.1)          |
| No   | 271 (98.9)       |
| <b>EQ5D utility index score (n = 262)</b>                                  | 0.490 (SD 0.260) |

<sup>a</sup>Shielding letter = letter to indicate clinical vulnerability requiring enhanced social distancing.

42  
43

#### 44 Assessment of rating scale function, item and scale refinement

45 Rating scale functioning was assessed by examining the relevant Winsteps output tables and item  
46 category probability curves (S1 Figure). Appraisal of category endorsement revealed the presence of  
47 floor effects and a positively skewed distribution. To address non-uniform category distribution, the 5-  
48 point rating scale was collapsed to either a dichotomous response or 4-point rating scale (where 0 =  
49 none, 1 = mild, 2 = moderate, 3 = severe). After rating scale adjustment, point-measure correlations  
50 were found to be positive (range = 0.23 - 0.92), categories were ordered, and category fit statistics  
51 were productive for measurement (Outfit MnSq < 2.0 logits) for all items. Category distribution  
52 remained positively skewed (mean skewness = 1.44, mean SE = 0.15, range = -2.30 - 11.64).  
53 Disordered thresholds were observed for 52 (39.7%) items. Threshold disordering is indicative of a  
54 category occupying a narrow interval on the rating scale and is only considered a cause for concern if  
55 category disordering is also observed [40]. In total, 35 items were removed (Fig 1 shows the process  
56 of item reduction and reasons for removal). The remaining 131 items were systematically grouped to  
57 construct scales that were clinically sensible and satisfied the Rasch model.

58

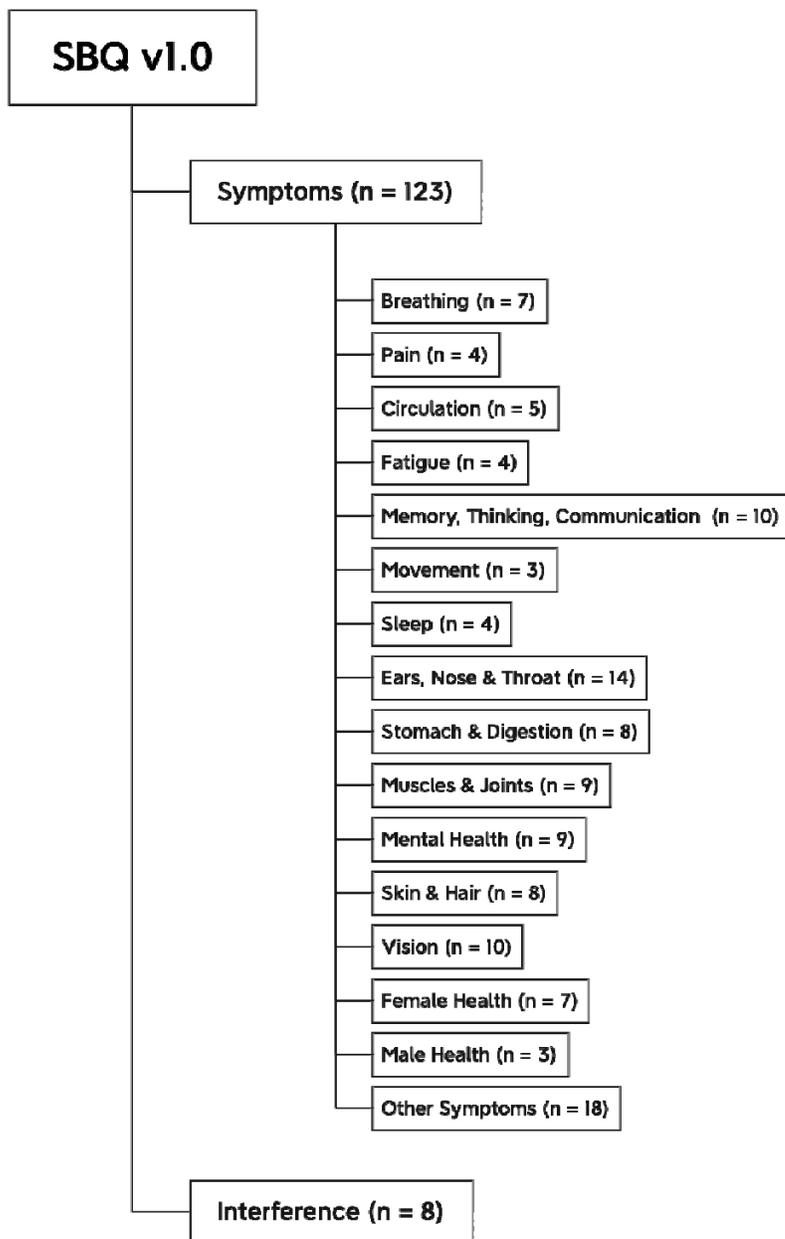
#### 59 Calibration of the final SBQ™-LC

60 Following response scale optimization and scale refinement, Rasch analyses were undertaken to  
61 report the psychometric properties of the final SBQ™-LC Version 1.0 (Table 3). The SBQ™-LC  
62 Version 1.0 is composed of 17 independent scales (Fig 2). All scales met the Rasch model criteria for  
63 unidimensionality and item fit. The first residual contrast values from principal component analyses of  
64 the residuals (PCAR) ranged from 1.46 to 2.03 eigenvalue units. No serious misfit was identified. Item  
65 infit MnSq values ranged from 0.67 to 1.32 logits and outfit MnSq values ranged from 0.44 to 1.53  
66 logits. Fourteen item pairs showed local item dependency with residual correlation values > 0.4  
67 (range = 0.44 – 0.88). In practical terms, a degree of local dependency is always observed in  
68 empirical data; therefore, it is necessary to consider the implications of item removal [51]. Despite  
69 evidence of local item dependence for eight scales, these items were retained to ensure  
70 comprehensive symptom coverage.

71 **Table 3.** Summary of Rasch analyses for the final SBQ™-LC (Version 1.0) (n = 274)

| Scale (Symptom domain)              | No items | Misfitting items | Persons with misfitting responses removed (%) | Mean person ability (logits) (SE) | PCA residual eigenvalue (1 <sup>st</sup> contrast) <sup>a</sup> | Dependent item pairs | Item separation | Item reliability | Person separation <sup>b</sup> | Person reliability <sup>c</sup> | Internal consistency reliability <sup>d</sup> |
|-------------------------------------|----------|------------------|---|-----------------------------------|---|----------------------|-----------------|------------------|--------------------------------|---------------------------------|---|
| Breathing                           | 7        | 0                | 31 (11.3)                                     | -1.95 (SE 0.99)                   | 1.72  | 0                    | 7.68            | 0.98             | 1.8                            | 0.76                            | 0.84  |
| Circulation                         | 5        | 0                | 17 (6.2)                                      | -1.16 (SE 0.98)                   | 1.58  | 0                    | 7.43            | 0.98             | 1.1                            | 0.55                            | 0.63  |
| Fatigue                             | 4        | 0                | 33 (12.0)                                     | 3.98 (SE 1.45)                    | 1.46  | 1                    | 1.34            | 0.64             | 1.78                           | 0.76                            | 0.91  |
| Memory, Thinking & Communication    | 10       | 0                | 12 (4.38)                                     | -1.04 (SE 0.54)                   | 1.76  | 0                    | 12.22           | 0.99             | 2.58                           | 0.87                            | 0.9   |
| Sleep                               | 4        | 0                | 31(11.3)                                      | -0.10 (SE 0.78)                   | 1.62  | 2                    | 13.44           | 0.99             | 1.67                           | 0.74                            | 0.56  |
| Movement                            | 3        | 0                | 31 (11.3)                                     | -6.27 (SE 2.06)                   | 1.57  | 3                    | 11.23           | 0.99             | 2.08                           | 0.81                            | 0.86  |
| Muscles & Joints                    | 9        | 0                | 23 (8.39)                                     | -0.85 (SE 0.54)                   | 1.68  | 0                    | 7.38            | 0.98             | 1.68                           | 0.73                            | 0.84  |
| Skin & Hair                         | 8        | 0                | 26 (9.49)                                     | -2.38 (SE 0.92)                   | 1.55  | 0                    | 4.6             | 0.95             | 0.71                           | 0.34                            | 0.68  |
| Eyes                                | 10       | 0                | 12 (4.38)                                     | -1.21 (SE 0.72)                   | 1.85  | 0                    | 4.28            | 0.95             | 1.08                           | 0.54                            | 0.72  |
| Ear, Nose & Throat                  | 14       | 0                | 11 (4.01)                                     | -1.25 (SE 0.49)                   | 1.96  | 1                    | 3.39            | 0.92             | 1.22                           | 0.6                             | 0.8   |
| Stomach & Digestion                 | 8        | 1                | 23 (8.39)                                     | -1.52 (SE 0.70)                   | 1.6   | 0                    | 6.19            | 0.97             | 0.95                           | 0.48                            | 0.7   |
| Mental Health & Wellbeing           | 9        | 0                | 25 (9.12)                                     | -0.70 (SE 0.52)                   | 1.66  | 0                    | 8.71            | 0.99             | 1.78                           | 0.76                            | 0.82  |
| Female Reproductive & Sexual Health | 7        | 0                | 27 (9.85)                                     | -2.07 (SE 1.05)                   | 1.99  | 2                    | 5.52            | 0.97             | 0.99                           | 0.5                             | 0.59  |
| Male Reproductive & Sexual Health   | 3        | 0                | 1 (3.03)                                      | -1.48 (SE 1.98)                   | 2.03  | 2                    | 2.27            | 0.84             | 0.79                           | 0.38                            | 0.6   |
| Pain                                | 4        | 0                | 25 (9.12)                                     | -1.53 (SE 1.05)                   | 1.75  | 2                    | 13.06           | 0.99             | 1.62                           | 0.72                            | 0.77  |
| Other symptoms                      | 18       | 0                | 22 (8.03)                                     | -1.53 (SE 0.43)                   | 1.67  | 1                    | 5.5             | 0.97             | 1.33                           | 0.64                            | 0.79  |
| Interference                        | 8        | 0                | 21 (7.66)                                     | 2.35 (SE 0.89)                    | 1.95  | 0                    | 11.21           | 0.99             | 2.13                           | 0.82                            | 0.89  |

72 <sup>a</sup> Unidimensionality: PCAR eigenvalue of first residual contrast < 2.0; Person separation: ≥1.50 acceptable, ≥ 2.00 good, ≥3.00 excellent; <sup>c</sup>Rasch-based reliability: r ≥ 0.70 acceptable, r ≥ 0.80 good, r ≥ 0.90 excellent; <sup>d</sup> Cronbach's α ≥ 0.70 acceptable



74

75 **Figure 2.** Conceptual framework of the final SBQ™-LC (Version 1.0)

76

77 Lastly, scale-to sample targeting and person reliability and separation were evaluated. Four scales  
78 had mean person ability values within  $\pm 1.0$  logits of mean item difficulty. Mean person ability ranged  
79 from -6.27 logits to 3.98 logits (SE range = 0.43 - 2.06). S2 Figure shows the item-person maps for  
80 the SBQ™-LC's scales. Person reliability ranged from 0.34 to 0.87 and person separation indices  
81 ranged from 0.71 – 2.58. Values for Cronbach's alpha, as evidence of internal consistency reliability,  
82 ranged from 0.56 to 0.91.

83

## 84 **DISCUSSION**

85 This study described the development and initial validation of the SBQ™-LC, a Rasch-developed,  
86 modular PRO instrument constructed to comprehensively assess symptom burden in Long COVID.

87 The SBQ™-LC was developed in accordance with international, consensus-based standards and  
88 regulatory guidance and may be used to evaluate the impact of interventions and to inform clinical  
89 care [14,22,52]. The findings from published systematic reviews were used to construct a conceptual  
90 framework and generate an initial item pool. Rigorous content validity testing provided evidence of the  
91 SBQ™-LC's relevance, comprehensiveness, comprehensibility, and acceptability. Rasch analysis  
92 permitted optimisation of the SBQ™-LC's items and response scales.

93

94 The SBQ™-LC was developed with the extensive involvement of adults with lived experience of Long  
95 COVID and is a strength of this study. Involvement of the target population is regarded as the “gold  
96 standard” in the development of PRO instruments. Patient involvement may be considered of  
97 particular importance in the context of Long COVID where there is a rapidly evolving evidence-base  
98 and affected individuals have reported experiences of stigma and a lack of acknowledgement from  
99 the medical community regarding the breadth and nature of their symptoms [9]. Involvement of adults  
100 with Long COVID in all study phases (i.e., development, refinement, and validation) ensured the  
101 patient perspective was embodied in the SBQ™-LC's items.

102

103 Rasch analysis of the SBQ™-LC guided item reduction and rating scale refinement, optimised  
104 measurement accuracy, and minimized respondent burden. These analyses showed several of the  
105 SBQ™-LC's scales to be off-target with evidence of low person reliability and separation. Poor scale-  
106 to sample targeting is indicative of the items within a scale failing to provide full coverage of person  
107 locations (i.e., range of symptom burden experienced by the sample) [53]. Negative mean person  
108 measures (>1.0 logits), floor effects, and positively skewed response category distributions suggested  
109 the SBQ™-LC may be targeting individuals with higher levels of symptom burden than the level of  
110 burden represented by the field test sample. Highly skewed samples and poor targeting can produce  
111 low person reliability coefficients even if an instrument is functioning as intended and provides a  
112 possible explanation for the low person reliability values observed in this study [53].

113

114 As a Rasch-developed instrument, the SBQ™-LC's ordinal raw scales may be converted to linear  
115 scales, with each 1-point change in a scale score being equidistant across the entire scale. Linear  
116 scores will enable the direct comparison of scores across the symptom domains represented by the  
117 SBQ™-LC's scales. This modular construction means researchers and clinicians need only select  
118 those scales required to provide targeted assessment of a particular symptom domain, reducing  
119 respondent burden further by removing the need to complete the SBQ™-LC in its entirety. Moreover,  
120 the Rasch model will make it possible to compare data from the SBQ™-LC with other PROs via co-  
121 calibration studies. As each item of a Rasch-derived scale functions independently from others on that  
122 scale, the SBQ™-LC can be adapted for administration as a computer adaptive test (CAT) [39,54]. A  
123 CAT is a test, administered via a computer, which adapts to the respondent's ability in real-time by  
124 selecting different questions from an item bank to provide a more accurate measure of the  
125 respondent's ability without the need to administer a large number of items [55]. CATs can reduce  
126 respondent burden, improve accuracy, and provide individualized assessment, instrument  
127 characteristics which are attractive when assessing a health condition with heterogeneous, relapsing  
128 and remitting symptoms such as Long COVID.

129

130 The burden of Long COVID on healthcare systems continues to grow as more individuals are infected  
131 by SARS-CoV-2 [56]. To meet this growing demand, services require cost-effective resources to  
132 support safe, effective clinical management. The use of the SBQ™-LC in the TLC study will provide  
133 early evidence of feasibility of the SBQ™-LC for use in remote patient monitoring [57]. A previous  
134 randomized controlled trial has shown remote symptom monitoring using PROs can result in fewer  
135 emergency department attendances, reduce hospital admissions, prompt earlier intervention, and  
136 improve patients' health-related quality of life [58]. If used in a clinical trial, symptom data collected  
137 remotely via the SBQ™-LC could provide valuable information on the safety, efficacy, and tolerability  
138 of new interventions for Long COVID [15]. If used within routine care, the SBQ™-LC has potential to  
139 facilitate patient-clinician conversations, guide treatment decision-making, and facilitate referrals to  
140 specialist services [58–60].

141

142 The use of social media to recruit participants meant it was not possible to confirm representativeness  
143 of the field test sample and is a limitation of this study. The relatively low completion rate suggested  
144 potential field test participants (i.e., possibly individuals experiencing higher levels of symptom  
145 burden, for example, symptoms such as fatigue or cognitive dysfunction) may have been deterred by  
146 the effort required to complete the SBQ™-LC or lacked sufficient incentive to participate. As such, the  
147 level of symptom burden captured in the field test may not reflect accurately the wider experience of  
148 Long COVID patients. Lastly, despite attempts to sample purposively for maximum variation, the  
149 demographics of the content validation study sample were highly skewed towards white (100%)  
150 females (76.9%). Further validation, undertaken as part of the TLC study which will prospectively  
151 target higher representation from ethnic minority groups, is planned and aims to confirm these  
152 findings. Further research is also needed to evaluate the SBQ™-LC using traditional psychometric  
153 indicators including test-retest reliability, construct validity, responsiveness, and measurement error.  
154 Studies to explore the feasibility and acceptability of the SBQ™-LC for use in clinical settings are also  
155 needed. The SBQ™-LC is currently available in UK English as an ePRO and in paper form. Linguistic  
156 and cross-cultural validation studies will ensure its generalizability outside of the UK and make  
157 translated versions of the SBQ™-LC available for use.

158

159 The presence of symptoms of COVID-19 persisting beyond the acute phase of infection in a  
160 significant number of patients represents an ongoing challenge for healthcare systems globally. High  
161 quality PRO instruments with robust psychometric properties are required to better understand the  
162 signs, symptoms, and underlying pathophysiology of Long COVID, to develop safe and effective  
163 interventions, and to meet the day-to-day needs of this growing patient group. The SBQ™-LC was  
164 developed as a comprehensive measure of Long COVID symptom burden. With promising  
165 psychometric properties, the SBQ™-LC may be suitable for use in Long COVID research studies and  
166 in the delivery of clinical care.

167

168 **Acknowledgements:**

169 The authors gratefully acknowledge the contributions of the clinicians who participated in the online  
170 survey. We thank Dr Anita Slade for their support with interpretation of the Rasch analyses.

171

172 The SBQ™-LC Version 1.0 is available under license. To request a copy, please e-mail:  
173 [tlcstudy@contacts.bham.ac.uk](mailto:tlcstudy@contacts.bham.ac.uk).

174

174 **Contributors:**

175 *Concept and design:* Calvert, Haroon, Hughes, Aiyegbusi, Frost, Davies

176 *Acquisition, analysis, or interpretation of data:* Hughes, Haroon, Calvert, Aiyegbusi, Matthews,  
177 Camaradou, Price, Ormerod, McMullan, Davies, Frost, McNamara  
178 *Drafting of the manuscript:* Hughes, Calvert, Haroon, Submaranian  
179 *Critical revision of the manuscript:* Hughes, Calvert, Haroon, Submaranian, McMullan, Aiyegbusi,  
180 Davies, Frost, McNamara, Matthews, Camaradou, Jackson, Turner, Price, Ormerod, Walker  
181 *Statistical analysis:* Hughes, Submaranian, Jackson  
182 *Obtained funding:* Haroon, Calvert, Hughes, Aiyegbusi, McMullan, Davies  
183 *Administration, technical, or material support:* Frost, Davies, Submaranian, Jackson, McNamara  
184 Walker

185  
186 **Competing interests:** SEH receives funding from the National Institute for Health Research (NIHR)  
187 Applied Research Collaboration (ARC) West Midlands, UK Research and Innovation (UKRI) and  
188 declares personal fees from Aparito Limited and Cochlear Limited outside the submitted work. **MJC** is  
189 Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of  
190 the Centre for the Centre for Patient Reported Outcomes Research and is a National Institute for  
191 Health Research (NIHR) Senior Investigator. MJC receives funding from the National Institute for  
192 Health Research (NIHR), UK Research and Innovation (UKRI), NIHR Birmingham Biomedical  
193 Research Centre, the NIHR Surgical Reconstruction and Microbiology Research Centre, NIHR ARC  
194 West Midlands, UK SPINE, European Regional Development Fund – Demand Hub and Health Data  
195 Research UK at the University of Birmingham and University Hospitals Birmingham NHS Foundation  
196 Trust, Innovate UK (part of UKRI), Macmillan Cancer Support, UCB Pharma, Janssen, GSK and  
197 Gilead. MJC has received personal fees from Astellas, Aparito Ltd, CIS Oncology, Takeda, Merck,  
198 Daiichi Sankyo, Glaukos, GSK and the Patient-Centered Outcomes Research Institute (PCORI)  
199 outside the submitted work. In addition, a family member owns shares in GSK. **OLA** receives funding  
200 from the NIHR Birmingham BRC, West Midlands, Birmingham, NIHR ARC West Midlands, UKRI,  
201 Health Foundation, Janssen, Gilead and GSK. He declares personal fees from Gilead Sciences Ltd,  
202 Merck and GSK outside the submitted work. **JC** is a lay member on the UK NICE COVID expert  
203 panel, a citizen partner to the COVID END Evidence Synthesis Network, PPI lead on the NIHR  
204 CICADA ME Study, patient representative at the EAN European Neurology Autonomic Nervous  
205 Systems Disorders Working Group, a member of the MRC/UKRI Advanced Pain Discovery Platform,  
206 and external board member of Plymouth Institute of Health. She also reports contracts with GSK and  
207 Medable. **CM** receives funding from NIHR SRMRC, UKRI, and declares personal fees from Aparito  
208 Ltd. **KM** is employed by the NIHR. **SH** receives funding from NIHR and UKRI. No other disclosures  
209 were reported.

210  
211 **Funding/Support:** This work is independent research jointly funded by the National Institute for  
212 Health Research (NIHR) and UK Research and Innovation (UKRI) (Therapies for Long COVID in non-  
213 hospitalized individuals: From symptoms, patient reported outcomes and immunology to targeted  
214 therapies (The TLC Study), (COV-LT-0013). The views expressed in this publication are those of the  
215 author(s) and not necessarily those of the NIHR, the Department of Health and Social Care, or UKRI.  
216

217 **Role of the Funders/Sponsors:** The funders had no role in the design and conduct of the study  
218 including: the collection, management, analysis, and interpretation of the data; preparation and review  
219 of the manuscript.  
220

221 **Data availability statement:** Data for this project are not currently available for access outside the  
222 TLC research team. The dataset may be shared when finalised but this will require an application  
223 to the data controllers. The data may then be released to a specific research team for a specific  
224 project dependent on the independent approvals being in place.  
225

## 226 REFERENCES

- 227 1. COVID-19 situation update worldwide, as of week 44, updated 11 November 2021. [cited 16 Nov  
228 2021]. Available from: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>  
229
- 230 2. National Institute for Health and Care Excellence (NICE). COVID-19 rapid guideline: managing  
231 the long-term effects of COVID-19. [cited 3 Dec 2021]. Available from:  
232 <https://app.magicapp.org/#/guideline/EQpzKn/section/n3vwoL>

- 233 3. Gapstur RL. Symptom Burden: A Concept Analysis and Implications for Oncology Nurses. *Oncol*  
234 *Nurs Forum*. 2007;34: 673–680. doi:10.1188/07.ONF.673-680. PMID: 17573326.
- 235 4. Aiyegbusi OL, Hughes SE, Turner G, Rivera SC, McMullan C, Chandan JS, Haroon S, Price G,  
236 Davies EH, Nirantharakumar K, Sapey E, Calvert MJ; TLC Study Group. Symptoms,  
237 complications and management of long COVID: a review. *J R Soc Med*. 2021 Sep;114(9):428-  
238 442. doi: 10.1177/01410768211032850. PMID: 34265229.
- 239 5. Cabrera Martimbianco AL, Pacheco RL, Bagattini ÂM, Riera R. Frequency, signs and symptoms,  
240 and criteria adopted for long COVID-19: A systematic review. *Int J Clin Pract*. 2021  
241 Oct;75(10):e14357. doi: 10.1111/ijcp.14357. PMID: 33977626.
- 242 6. CDC. Post-COVID Conditions. In: Centers for Disease Control and Prevention [Internet]. 11 Feb  
243 2020 [cited 7 Dec 2021]. Available from: [https://www.cdc.gov/coronavirus/2019-ncov/long-term-](https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html)  
244 [effects/index.html](https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html)
- 245 7. Amin-Chowdhury Z, Harris RJ, Aiano F, Zavala M, Bertran M, Borrow R, et al. Characterising  
246 long COVID more than 6 months after acute infection in adults; prospective longitudinal cohort  
247 study, England. *medRxiv [Preprint]*. 2021 Mar [cited 2022 Jan 14]. Available from:  
248 <https://www.medrxiv.org/content/10.1101/2021.12.20.21268098v1>  
249 doi:10.1101/2021.03.18.21253633
- 250 8. Parker AM, Brigham E, Connolly B, McPeake J, Agranovich AV, Kenes MT, Casey K, Reynolds  
251 C, Schmidt KFR, Kim SY, Kaplin A, Sevin CM, Brodsky MB, Turnbull AE. Addressing the post-  
252 acute sequelae of SARS-CoV-2 infection: a multidisciplinary model of care. *Lancet Respir Med*.  
253 2021 Nov;9(11):1328-1341. doi: 10.1016/S2213-2600(21)00385-4. PMID: 34678213.
- 254 9. Ladds E, Rushforth A, Wieringa S, Taylor S, Rayner C, Husain L, Greenhalgh T. Persistent  
255 symptoms after Covid-19: qualitative study of 114 "long Covid" patients and draft quality  
256 principles for services. *BMC Health Serv Res*. 2020 Dec 20;20(1):1144. doi: 10.1186/s12913-  
257 020-06001-y. PMID: 33342437.
- 258 10. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence,  
259 and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618  
260 survivors of COVID-19. *PLoS Med*. 2021 Sep 28;18(9):e1003773. doi:  
261 10.1371/journal.pmed.1003773. PMID: 34582441.
- 262 11. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and  
263 predictors of long COVID. *Nat Med*. 2021;27(4): 626–631. doi:10.1038/s41591-021-01292-y.  
264 PMID: 33692530.
- 265 12. Nasserie T, Hittle M, Goodman SN. Assessment of the Frequency and Variety of Persistent  
266 Symptoms Among Patients With COVID-19: A Systematic Review. *JAMA Netw Open*. 2021  
267 May 3;4(5):e2111417. doi: 10.1001/jamanetworkopen.2021.11417. PMID: 34037731.
- 268 13. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Author  
269 Correction: Attributes and predictors of long COVID. *Nat Med*. 2021 Jun;27(6):1116. doi:  
270 10.1038/s41591-021-01361-2. Erratum for: *Nat Med*. 2021 Apr;27(4):626-631. PMID: 34045738
- 271 14. FDA. Guidance for Industry Use in Medical Product Development to Support Labeling Claims  
272 Guidance for Industry. *Clinical/Medical Federal Register*. 2009; 1–39. doi:10.1111/j.1524-  
273 4733.2009.00609.x
- 274 15. Aiyegbusi OL, Calvert MJ. Patient-reported outcomes: central to the management of COVID-19.  
275 *Lancet*. 2020 Aug 22;396(10250):531. doi: 10.1016/S0140-6736(20)31724-4. Epub 2020 Aug  
276 10. PMID: 32791038.
- 277 16. Berry P. Use patient reported outcome measures (PROMs) in treatment of long covid. *BMJ*. 2021  
278 May 18;373:n1260. doi: 10.1136/bmj.n1260. PMID: 34006526.

- 279 17. Calvert M, Brundage M, Jacobsen PB, Schünemann HJ, Efficace F. The CONSORT Patient-  
280 Reported Outcome (PRO) extension: implications for clinical trials and practice. *Health Qual Life*  
281 *Outcomes*. 2013 Oct 29;11:184. doi: 10.1186/1477-7525-11-184. PMID: 24168680.
- 282 18. Tran VT, Riveros C, Cleprier B, Desvarieux M, Collet C, Yordanov Y, et al. Development and  
283 validation of the long covid symptom and impact tools, a set of patient-reported instruments  
284 constructed from patients' lived experience. *Clin Infect Dis*. 2021 Apr 29:ciab352. doi:  
285 10.1093/cid/ciab352. Epub ahead of print. PMID: 33912905.
- 286 19. O'Connor RJ, Preston N, Parkin A, Makower S, Ross D, Gee J, et al. The COVID-19 Yorkshire  
287 Rehabilitation Scale (C19-YRS): Application and psychometric analysis in a post-COVID-19  
288 syndrome cohort. *J Med Virol*. 2021 Oct 21:10.1002/jmv.27415. doi: 10.1002/jmv.27415. PMID:  
289 34676578.
- 290 20. Alwan NA. The teachings of Long COVID. *Commun Med*. 2021;1: 1–3. doi:10.1038/s43856-021-  
291 00016-0
- 292 21. Callard F, Perego E. How and why patients made Long Covid. *Soc Sci Med*. 2021  
293 Jan;268:113426. doi: 10.1016/j.socscimed.2020.113426. PMID: 33199035.
- 294 22. FDA. Principles for Selecting, Developing, Modifying, and Adapting Patient-Reported Outcome  
295 Instruments for Use in Medical Device Evaluation: Draft Guidance for Industry and Food and  
296 Drug Administration Staff, And Other Stakeholders. 2020 Aug [cited 2022 Jan 14]. Available  
297 from: [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-selecting-developing-modifying-and-adapting-patient-reported-outcome-instruments-use)  
298 [selecting-developing-modifying-and-adapting-patient-reported-outcome-instruments-use](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-selecting-developing-modifying-and-adapting-patient-reported-outcome-instruments-use)
- 299 23. Haroon S, Nirantharakumar K, Hughes SE, Subramanian A, Aiyegbusi OL, Davies EH, et al.  
300 Therapies for Long COVID in non-hospitalised individuals - from symptoms, patient-reported  
301 outcomes, and immunology to targeted therapies (The TLC Study): Study protocol. *medRxiv*  
302 [Preprint]. 2021 [cited 14 Jan 2022]. Available from:  
303 <https://www.medrxiv.org/content/10.1101/2021.12.20.21268098v1>  
304 doi:10.1101/2021.12.20.21268098
- 305 24. Linacre JM. Sample Size and Item Calibration or Person Measure Stability. *Rasch Measurement*  
306 *Transactions*. 1994;7: 328.[Cited 2022 Jan 14]. Available from:  
307 <https://www.rasch.org/rmt/rmt74m.htm>
- 308 25. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al.  
309 More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep*.  
310 2021 Aug 9;11(1):16144. doi: 10.1038/s41598-021-95565-8. PMID: 34373540.
- 311 26. de Haes JC, van Knippenberg FC, Neijt JP. Measuring psychological and physical distress in  
312 cancer patients: structure and application of the Rotterdam Symptom Checklist. *Br J Cancer*.  
313 1990 Dec;62(6):1034-8. doi: 10.1038/bjc.1990.434. PMID: 2257209.
- 314 27. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the  
315 National Cancer Institute's patient-reported outcomes version of the common terminology  
316 criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst*. 2014 Sep 29;106(9):dju244. doi:  
317 10.1093/jnci/dju244. PMID: 25265940.
- 318 28. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom  
319 distress in cancer patients: the M.D. Anderson Symptom Inventory. *Cancer*. 2000 Oct  
320 1;89(7):1634-46. doi: 10.1002/1097-0142(20001001)89:7<1634::aid-cnrcr29>3.0.co;2-v. PMID:  
321 11013380.
- 322 29. Powers JH, Guerrero ML, Leidy NK, Fairchok MP, Rosenberg A, Hernández A, et al.  
323 Development of the Flu-PRO: a patient-reported outcome (PRO) instrument to evaluate  
324 symptoms of influenza. *BMC Infect Dis*. 2016 Jan 5;16:1. doi: 10.1186/s12879-015-1330-0.  
325 PMID: 26729246.

- 326 30. Evans RA, McAuley H, Harrison EM, Shikotra A, Singapuri A, Sereno M, et al. Physical,  
327 cognitive, and mental health impacts of COVID-19 after hospitalisation (PHOSP-COVID): a UK  
328 multicentre, prospective cohort study. *Lancet Respir Med*. 2021;9: 1275–1287.  
329 doi:10.1016/S2213-2600(21)00383-0. PMID: 34627560.
- 330 31. Smart Survey. Available from: <https://www.smartsurvey.co.uk/>
- 331 32. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3: 77–101.
- 332 33. Willis GB. *Analysis of the Cognitive Interview in Questionnaire Design*. Oxford: Oxford University  
333 Press; 2015.
- 334 34. Aparito Limited [Cited 2022 Jan 14]. Available from: <https://www.aparito.com/>
- 335 35. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielson  
336 and Lydiche; 1960. doi:10.1016/S0019-9958(61)80061-2
- 337 36. Bond TG, Yan Z, Heene M. *Applying the Rasch Model: Fundamental Measurement in the Human  
338 Sciences*. 4th edition. Routledge; 2020.
- 339 37. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: The  
340 role of new psychometric methods. *Health Technol Assess*. 2009;13. doi:10.3310/hta13120.  
341 PMID: 19216837.
- 342 38. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and  
343 application of quality-of-life instruments. *Value Health*. 2004;7: S22–S26. doi:10.1111/j.1524-  
344 4733.2004.7s106.x. PMID:15367240.
- 345 39. Linacre JM. *A User's Guide to WINSTEPS MINISTEPS Rasch-Model Computer Programs*. 2020.  
346 doi:ISBN 0-941938-03-4
- 347 40. Mousavi A, Cui Y. The Effect of Person Misfit on Item Parameter Estimation and Classification  
348 Accuracy: A Simulation Study. *Education Sciences*. 2020;10: 324.  
349 doi:10.3390/educsci10110324
- 350 41. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why  
351 use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis  
352 Rheum*. 2007. doi:10.1002/art.23108. PMID: 18050173.
- 353 42. Linacre J. Optimizing Rating Scale Category Effectiveness. *J Appl Meas*. 2002;3(1): 85–106.  
354 PMID:11997586.
- 355 43. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome scale:  
356 The stroke impact scale. *Arch Phys Med Rehabil*. 2003;84: 950–963. doi:10.1016/S0003-  
357 9993(03)00035-2. PMID: 12881816.
- 358 44. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A practical guide to their  
359 development and use*. Health Measurement Scales. Oxford University Press; 2015.
- 360 45. Flesch R. A new readability yardstick. *J Applied Psychol*. 1948;32(3): 221–233.  
361 doi:10.1037/h0057532. PMID: 18867058.
- 362 46. McLaughlin GH. SMOG grading: A new readability formula. *Journal of Reading*. 1969;12: 639–  
363 646.
- 364 47. Readable. Available: [www.readable.io](http://www.readable.io)
- 365 48. Weiss BD. *Help patients understand: Manual for clinicians*. American Medical Association  
366 Foundation and American Medical Association; 2007 May [cited 2022 Jan 14]. Available from:

- 367 [https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-](https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-clinicians-2nd-ed)  
368 [clinicians-2nd-ed](https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-clinicians-2nd-ed)
- 369 49. NHS Digital Service Manual. Use a readability tool to prioritise content - NHS digital service  
370 manual. In: nhs.uk [Internet]. Jul 2019 [cited 22 Jul 2020]. Available from: [https://service-](https://service-manual.nhs.uk)  
371 [manual.nhs.uk](https://service-manual.nhs.uk)
- 372 50. Local Dependency and Rasch Measures. [cited 10 Dec 2021]. Available from:  
373 <https://www.rasch.org/rmt/rmt213b.htm>
- 374 51. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN  
375 Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life*  
376 *Res.* 2018 May;27(5):1171-1179. doi: 10.1007/s11136-017-1765-4. PMID: 29260445;
- 377 52. Separation, Reliability and Skewed Distributions: Statistically Different Levels of Performance.  
378 [cited 6 Dec 2021]. Available from: <https://www.rasch.org/rmt/rmt144k.htm>
- 379 53. Harrison CJ, Geerards D, Ottenhof MJ, Klassen AF, Riff KWW, Swan MC, et al. Computerised  
380 adaptive testing accurately predicts CLEFT-Q scores by selecting fewer, more patient-focused  
381 questions. *J Plast Reconstr Aesthet Surg.* 2019. doi:10.1016/j.bjps.2019.05.039. PMID:  
382 31358447.
- 383 54. Computer Adaptive Testing: Background, benefits and case study of a large-scale national testing  
384 programme. In: *Surpass, Powering Assessment* [Internet]. 11 Dec 2019 [cited 5 Dec 2021].  
385 Available from: [https://surpass.com/news/2019/computer-adaptive-testing-background-benefits-](https://surpass.com/news/2019/computer-adaptive-testing-background-benefits-and-case-study-of-a-large-scale-national-testing-programme/)  
386 [and-case-study-of-a-large-scale-national-testing-programme/](https://surpass.com/news/2019/computer-adaptive-testing-background-benefits-and-case-study-of-a-large-scale-national-testing-programme/)
- 387 55. Menges D, Ballouz T, Anagnostopoulos A, Aschmann HE, Domenghino A, Fehr JS, et al. Burden  
388 of post-COVID-19 syndrome and implications for healthcare service planning: A population-  
389 based cohort study. *PLOS ONE.* 2021;16: e0254523. doi:10.1371/journal.pone.0254523. PMID:  
390 34252157
- 391 56. Vindrola-Padros C, Sidhu MS, Georghiou T, Sherlaw-Johnson C, Singh KE, Tomini SM, et al.  
392 The implementation of remote home monitoring models during the COVID-19 pandemic in  
393 England. *EClinicalMedicine.* 2021;34. doi:10.1016/j.eclinm.2021.100799. PMID: 33817610.
- 394 57. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with  
395 patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *J Clin*  
396 *Oncol.* 2016;34: 557–565. doi:10.1200/JCO.2015.63.0830. PMID: 26644527.
- 397 58. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient  
398 reported outcome measures (PROMs) support clinician-patient communication and patient  
399 care? A realist synthesis. *J Patient Rep Outcomes.* 2018;2. doi:10.1186/s41687-018-0061-6.  
400 PMID: 30294712.
- 401 59. Calvert M, Kyte D, Price G, Valderas JM, Henrik Hjollund N. Maximising the impact of patient  
402 reported outcome assessment for patients and society. *BMJ.* 2019;364. doi:10.1136/bmj.k5267.  
403 PMID: 30679170.
- 404

405

## SUPPORTING INFORMATION CAPTIONS

406 **S1 Table. Themes and exemplar quotations from content validation of the SBQ™-LC item pool**

407 **(cognitive debriefing and online survey)**

408

409 **S1 Figure. Item category probability curves for the final SBQ™-LC (Version 1.0)**

410 Each graph, the y-axis represents the expected probability of endorsement of any given category

411 when a person responds to the item. The x-axis represents the person ability relative to the item

412 difficulty, with origins set to 0. The scale is measured in logits. Locations to the right of '0' represent

413 greater and greater levels of symptom burden. Moving to the left of '0' locations represent lower levels

414 of symptom burden. Categories should advance in ascending order from left to right along the x-axis

415 and each have a distinct peak (indicating that the category is the most probable (modal) category at

416 that point on the latent variable). The crossover points between two curves are the equal probability

417 points of thresholds.

418

419 **S2 Figure. Item-person maps for the 17 scales of the final SBQ™-LC (Version 1.0)**

420 For each scale, the person-item map displays the location of person abilities and item difficulties

421 respectively along the same latent dimension (y-axis).