

<https://helda.helsinki.fi>

Reliability and validity of the Finnish version of the prosthesis evaluation questionnaire

Repo, Jussi Petteri

2018

Repo , J P , Piitulainen , K , Häkkinen , A , Roine , R P , Kautiainen , H , Becker , P & Tukiainen , E J 2018 , ' Reliability and validity of the Finnish version of the prosthesis evaluation questionnaire ' , Disability and Rehabilitation , vol. 40 , no. 17 , pp. 2081-2087 . <https://doi.org/10.1080/09638288.2017.1323032>

<http://hdl.handle.net/10138/311586>

<https://doi.org/10.1080/09638288.2017.1323032>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

1 **Reliability and validity of the Finnish version of the Prosthesis Evaluation**

2 **Questionnaire**

3

4 **Abstract**

5 **BACKGROUND** Thus far there have been no specific patient-reported instruments in
6 Finnish for health-related quality of life assessment after major lower extremity
7 amputation and successful prosthesis fitting.

8 **METHODS** The Prosthesis Evaluation Questionnaire was translated and cross-
9 culturally adapted to Finnish. Participants completed a questionnaire package including
10 the Finnish version of the Prosthesis Evaluation Questionnaire and the 15D health-
11 related quality of life instrument. Scales ($n = 10$) were tested for internal consistency,
12 floor-ceiling effect, and reproducibility for which participants completed the Prosthesis
13 Evaluation Questionnaire twice within a 2-week interval. Validity was tested by
14 estimating the correlation between the 15D index and the scales. The authors included
15 122 participants who had completed the questionnaire on two separate occasions in the
16 final analysis.

17 **RESULTS** Mean scale scores of the 10 scales varied from 52 to 83. Cronbach alphas
18 ranged from 0.67 to 0.96. The total score showed no floor-ceiling effect.
19 Reproducibility of the scales was good (intraclass correlation coefficient, 0.78-0.87;
20 coefficient of repeatability, 19-36). Significant correlations were observed between the
21 15D index and the scales for Ambulation, Social Burden, Usefulness and Well-being.

22 **CONCLUSIONS** This study provided evidence of the reliability and validity of the
23 Finnish version of the Prosthesis Evaluation Questionnaire in assessing the health-related
24 quality of life among major lower extremity amputated patients who have been fitted
25 with prosthesis.

26 **Key Words:** Rehabilitation; Prosthesis; Validity; Reliability; Amputation;
27 Psychometrics

28

29 **Introduction**

30 Assessing rehabilitation effectiveness with high quality patient-reported outcome
31 instruments makes it possible to obtain an amputee-centered experience in a relevant
32 way [1]. Generic instruments that are designed to obtain information from a broad
33 variety of health parameters may not be specific enough to measure the specific
34 problems encountered by amputees. Thus, the Prosthesis Profile of the Amputee
35 questionnaire was introduced in 1994 to provide a tool for lower-extremity amputee-
36 specific assessment [2]. Qualitative studies may provide deep insight into patients'
37 biopsychosocial perspectives that otherwise would be hard to obtain. However,
38 quantitative data obtained from patient-reported instruments can provide accurate and
39 reliable outcomes that can be statistically analyzed for the assessment of effectiveness
40 of different methods of surgical approaches or rehabilitation.

41

42 The rehabilitation assessment further evolved towards emphasizing the impact of
43 rehabilitation on the health-related quality of life (HRQoL), when Legro et al.
44 developed and validated the English version of the Prosthesis Evaluation Questionnaire
45 (PEQ) [3]. The PEQ is an amputee-specific quality of life instrument that can be used to
46 assess the HRQoL of lower-extremity prosthesis users. It has been further
47 psychometrically investigated and validated after being translated into several other
48 languages [4-8]. Furthermore, the PEQ has been used in a great variety of studies [9].

49

50 There has hitherto been no validated lower-extremity amputee-specific patient-reported

51 outcome instrument in Finnish. The authors aimed, therefore, to transculturally adapt
52 the English PEQ into a Finnish version, which was then tested for reliability and
53 validity among patients who have undergone major lower extremity amputation and
54 have rehabilitated to prosthesis users.

55

56 **Methods**

57 *Ethical considerations and participants*

58 The Ethics Committee of the Helsinki University Hospital approved the study. The
59 authors included in the study patients, who had undergone major lower extremity
60 amputation, were at least 18 years old, had full ability to understand written Finnish and
61 had rehabilitated to prosthesis users in the Helsinki and Uusimaa Hospital District or the
62 Central Finland Health Care District, Finland. The participants provided their written
63 consent according to the Helsinki Declaration. The authors approached by mail 597
64 consecutive patients who had undergone major lower extremity amputation and had
65 successful prosthesis fitting.

66

67 *Translation and adaptation*

68 The authors contacted the developer of the PEQ to obtain permission to use the English
69 language questionnaire. The translation and adaptation process adhered to the
70 International Society for Pharmacoeconomics and Outcomes Research guidelines [10].

71

72 Two native Finnish-speaking translators who were professionals in the field of
73 rehabilitation and fluent in English produced a forward-translation independently of
74 each other. Differences encountered between the two forward translations were
75 discussed by the steering group who then synthesized one forward-translation. A back-

76 translation was produced by an English language-expert who was fluent in Finnish and
77 familiar with the Finnish culture and translating patient-reported outcome instruments
78 but unfamiliar with the current instrument. A back-translation panel consisting of all
79 three translators reviewed the translation drafts and compared them to the original
80 English version and provided a written report. In addition to this a language expert of
81 the Finnish Medical Society Duodecim was consulted when translation problems were
82 encountered. A multidisciplinary committee reviewed each part of the translation
83 processes separately.

84

85 The pre-final version underwent pre-testing together with cognitive debriefing among
86 14 Finnish patients who had undergone transtibial amputation and who were transtibial
87 prosthesis users. The cognitive debriefing followed the European Organisation for
88 Research and Treatment of Cancer (EORTC) guidelines [11] to identify any offensive
89 content, problems with understandability, cultural relevance, difficulties in answering or
90 in interpretation of the questions and whether the participants would ask any question
91 differently. In the last phase, the multidisciplinary committee reviewed the pre-testing
92 outcomes and interview reports. The final version was introduced and was then
93 proofread by the language expert of the Finnish Medical Society Duodecim
94 (Supplementary file).

95

96 ***Instruments***

97 ***Prosthesis Evaluation Questionnaire.*** The lower extremity amputee-specific PEQ is a
98 valid, comprehensive instrument comprising 82 items with seven different main themes.
99 The items refer to the preceding four weeks. The PEQ also contains items with

100 checkboxes for assessing topics such as Satisfaction, Pain, Transfers, Prosthesis care
101 and Self-efficacy. These items are scored individually.

102

103 The PEQ can be separated into 10 validated scales: Ambulation, Appearance,
104 Frustration, Perceived Response, Social Burden, Utility, Residual Limb Health, Sounds,
105 Transfers and Well-being [3]. The items are completed on a visual analogue scale (0-
106 100 mm; from worst to best). The total scores for each scale are calculated through the
107 arithmetic mean of all items of the scale.

108

109 ***15D instrument.*** The 15D is a valid, generic HRQoL instrument containing 15
110 dimensions: moving, seeing, hearing, breathing, sleeping, eating, speech, excretion,
111 usual activities, mental function, discomfort and symptoms, depression, distress, vitality
112 and sexual activity [12]. Respondents choose one of the five levels in each dimension
113 that best describes their current state of health (1-5; best to worst possible score). The
114 15D produces both a HRQoL profile and a single index score that represents the overall
115 HRQoL. The single index score ranges from 0 to 1, with 0 equivalent to being dead and
116 1 being in the best imaginable HRQoL state. Reproducibility and the minimum
117 important change of the 15D have been reported to be 0.90 and 0.015, respectively
118 [13,14].

119

120 ***Sociodemographic and clinical questionnaire.*** The authors obtained information on
121 participants' age, sex, cause for amputation, comorbidities, amputation level
122 (disarticulation amputation was considered as above-knee amputation), time since
123 amputation, and beginning of the prosthesis use. In addition, a visual analogue scale on
124 a 0 to 100 mm scale (0-100 mm; best to worst) was used for measuring participants'

125 self-reported general health and pain during the preceding week. The NRS is another
126 instruments as it is a segmented numeric version of the visual analog scale (VAS) in
127 which a respondent selects a whole number (0–10 integers) that best reflects the
128 intensity of their pain. The visual analog scale, which the authors used, is a widely
129 accepted measure and validated for pain assessment [15].

130

131 *PEQ validation course and reproducibility setting*

132 In addition to the pre-information form, the authors included the following instruments
133 in the first questionnaire package: the Finnish PEQ, the 15D and the general health and
134 pain visual analogue scale questions. Participants returned the completed questionnaires
135 together with the signed informed consent. Potential participants who did not return the
136 first questionnaire set within a week received a reminder letter. After the participants
137 had completed the first questionnaire, the authors mailed them the PEQ instrument a
138 second time along with a survey. The purpose of both was to ascertain whether the
139 patients' health status had changed between completing the first round of
140 questionnaires. The authors included participants who had completed the PEQ twice in
141 the final analyses.

142

143 *Statistics*

144 The authors present the data as means with standard deviations (SD), medians with
145 interquartile ranges (IQR), 95% confidence intervals (95% CI), or as counts with
146 percentages or ranges. The scale completion rate is provided to illustrate the percentage
147 of missing items in the analysis. Predefined hypotheses were placed based on the
148 existing literature or general presumptions [table 1].

149

150 A one-way random-effects model with absolute agreement was used to measure relative
151 reliability or intraclass correlation coefficient. The intraclass correlation coefficient
152 value was classified according to Cicchetti et al. as poor (< 0.40), fair ($0.40-0.59$), good
153 ($0.60-0.74$) or excellent ($0.75-1.00$) [16].

154

155 The internal consistency was estimated by calculating Cronbach's alpha [17] with
156 bootstrapped 95% CIs.

157

158 The coefficient of repeatability expressed the expected maximum size of 95% of the
159 absolute differences between paired observations. The 95% CI was obtained by bias
160 corrected and accelerated bootstrapping (5000 replications).

161

162 The Pearson method served to calculate the correlation coefficients. Statistical
163 significance in the correlation coefficient was set at $p < 0.05$ and calculated using Sidak-
164 adjusted probabilities. Bias-corrected bootstrapping was used to obtain the confidence
165 intervals for the mean changes between the two measurements and reproducibility.

166

167 The authors used linear regression analyses to identify the appropriate predictors of the
168 15D age- and gender-standardized regression coefficients Beta (β). The β -value is a
169 measure of how strongly each predictor variable influences the criterion (dependent)
170 variable. The β was measured in units of standard deviation. Cohen's standard for β -
171 values above 0.10, 0.30 and 0.50 represent small, moderate and large relationships,
172 respectively.

173

174

175 **Results**

176 Of the 167 participants (response rate, 28%), who returned the questionnaires together
177 with their signed written consent, a total of 122 patients (73%) had completed both the
178 first and the second questionnaires and were included in the study. The participants'
179 ages ranged from 19 to 93 [table 2]. The most common indication for primary major
180 lower-extremity amputation was vascular disease (29.5%). Thirty-six percent ($n = 44$)
181 of participants reported having no comorbidities [table 2]. The time from amputation to
182 completion of the outcome measures varied from four months to 69 years. Fifty percent
183 of the participants had undergone amputation less than five years earlier.

184

185 ***Translation and adaptation***

186 Minor linguistic differences were noted between the two forward translations. A back-
187 translation panel review revealed no major problems between the back-translation and
188 the original English version. The multidisciplinary committee required that “rate the
189 weight of your prosthesis” in item 1C be changed to “evaluate the weight of your
190 prosthesis” in order to improve the Finnish. Item 1N required amending “prosthesis
191 cover” to “cosmetic surface” which is preferred in Finnish. In item 1Q the word
192 “stump” was added for clarification. In the Finnish language, the word “stump” is well
193 accepted to describe the distal end of an amputated limb. Translation of the words: ”
194 “shooting”, “searing”, “stabbing”, “sharp”, “ache” in the “Group 2” of the PEQ
195 instrument required the help of the language expert to find suitable matches in Finnish.
196 The pre-testing and participants’ cognitive debriefing gave no reason for changes.

197

198

199

200 **Reliability**

201 **Floor-ceiling effect.** The PEQ showed no floor-effect (0 score) on the total score. Nine
202 of the scales had no floor effect. Altogether 1% had the lowest score in Ambulation
203 scale. A ceiling effect of one to five percent was found in five of the scales [table 3].
204 The highest ceiling effect was strongest in the Perceived Responses scale (5%).

205

206 **Internal consistency.** Cronbach's alpha for the 10 scales revealed an internal
207 consistency ranging from 0.67 (Appearance) to 0.96 (Ambulation) [table 3].

208

209 **Reproducibility.** The mean value (SD) of the PEQ subscales at measurement one was
210 65.1 (23.7) (table 4). The mean change between the two measurement times ranged
211 from 0.0 to 2.1 in the separate scales. All scales had good reproducibility [table 4]. The
212 coefficient of repeatability ranged from 19 for Usefulness to 36 for the Frustration
213 scales [table 4].

214

215 **Validity**

216 **Convergent validity.** Pearson correlation coefficients between the PEQ scales and age
217 were low (range, -0.28 to 0.15) [table 5]. The correlation of the PEQ scale scores with
218 time since prosthetization was also poor. Strong correlation was found between general
219 pain or general health and Usefulness, Ambulation, Transfers, Perceived responses,
220 Social Burden and Well-Being scales.

221

222 Strong correlation was found between the 15D index and the scales of Ambulation
223 Social burden, Transfers, Usefulness and Well-being [Figure 1].

224

225 **Discussion**

226 The authors successfully produced a Finnish PEQ instrument and evaluated its
227 psychometric properties. To the authors' knowledge this study has the largest study
228 population to assess the psychometrics of the PEQ. The psychometric analyses showed
229 evidence of good reproducibility and validity for the Finnish PEQ. The Finnish version
230 of the PEQ instrument can now be used to assess the effectiveness of different
231 amputation techniques, stump reconstruction methods, and rehabilitation after
232 successful prosthesis fitting.

233

234 *Translation and adaptation*

235 The translation and cross-cultural adaptation process adhered rigorously to the
236 International Society for Pharmacoeconomics and Outcomes Research guidelines [10].
237 All the discrepancies and changes made during the translation phases were meticulously
238 recorded in written reports. One previous translation report addressed the linguistic or
239 cultural problems encountered during the translation process [5]. The authors found
240 that adjustments were required to adjust for linguistic differences between the Finnish
241 version of the PEQ and the original English version.

242

243 In the Arabic translation of the PEQ, the authors found the word “phantom” could be
244 interpreted as a “ghost sensation” among the Saudi people [5]. The word “phantom”
245 does not have a negative connotation in Finnish nor is it linked to ghosts. The
246 identification of items in the Arabic version was changed to match the group number
247 rather than the page number as in the original English version [3,5]. The Finnish version
248 also uses the group numbers to identify the items. The new numbering of items should
249 be taken into consideration when using the Finnish PEQ.

250 ***Reliability***

251 A floor-ceiling effect of less than 15% is considered acceptable [18]. Reliability testing
252 for the PEQ by Legro et al. found a floor effect of 22% in the scales of Frustration and a
253 ceiling effect of 25% in the Transfers scale in a similar study population to that of the
254 present study [3]. No explanation for this was provided by the Legro group. It could be
255 hypothesized however that the ceiling effect was a consequence of the answers of those
256 participants who had been amputated 9 to 28 years before assessment took place as they
257 received the highest scores in the Transfers scale [3]. Other validation studies of PEQ
258 did not report floor-ceiling values [5-8]. In the present study, five percent of participants
259 received the maximum score in the Perceived Responses scale. Not a single participant
260 reported the maximum scores in the Transfers scale. The PEQ scales of the Finnish
261 version seemed to have no floor or ceiling effect based on the present study's findings.
262 Thus, the present analysis provided evidence that it is somewhat unlikely that the PEQ
263 would yield inaccurate maximum scores.

264

265 The internal consistency of the original English PEQ varies between 0.67 and 0.89 in
266 the 10 scales [3]. Cronbach's alphas between 0.67 and 0.96 were noted in the present
267 study. According to the literature, Cronbach's alpha of 0.8 or more is considered
268 sufficient [19]. In the present study four of the 10 subscales were slightly lower than the
269 proposed benchmark, but these values can be considered acceptable. Benavent et al.
270 found poor internal consistency in the scales of Appearance and Residual Limb (0.37
271 and 0.15, respectively) [8]. Cronbach's alpha varied in the remaining scales between
272 0.55 and 0.93 in that study [8]. Other studies have reported the internal consistency of
273 Appearance and Residual Limb Health of 0.73-0.77 and 0.77-0.80, respectively [3,6,7].
274 The results of the present study were similar to those the previous studies [3,6,7] as the

275 internal consistency of the Appearance scale was 0.79 and that of the Residual Limb
276 was 0.67. Internal consistency of the other eight scales were also mainly in concordance
277 with previously published literature [3,6-8]. The internal consistency in the present
278 study was below 0.9 in all scales, indicating that there was no item repetition [20].
279 The authors assessed reproducibility after a mean interval of two weeks. The
280 participants' health was stable in the interim period. The optimal interim time between
281 the two assessments has previously been placed at two weeks in assessment of the
282 reproducibility in situations where there is no acute change in the participants' health
283 [21]. According to the classification by Cicchetti et al. [14], all scales used in the
284 Finnish PEQ had excellent intraclass correlation coefficient values (0.78-0.87). Conrad
285 and colleagues reported intraclass correlation coefficient values that ranged from good
286 (0.65, Well-being) to excellent (0.92, Ambulation) between the scales in the Brazilian
287 Portuguese version of the PEQ [6]. However, the Conrad group reported on a smaller
288 study population that consisted only of 65 patients who had undergone major lower-
289 extremity amputation [6]. The authors also calculated the coefficient of repeatability for
290 the PEQ scales. The coefficient of repeatability can be used to obtain the value for
291 absolute reliability, the expected maximum size of 95% of the absolute differences
292 between paired observations. The present study reflects the good reproducibility of the
293 PEQ instrument scales. The authors found that the coefficient of repeatability ranged
294 from 19 to 36 between the different scales in the present study. The alternative of
295 calculating the coefficient of repeatability values may be more accurate compared to the
296 standard error of measurement as it takes into account both random and systematic
297 errors [22].
298
299

300 **Validity**

301 Age has previously been reported to correlate with Residual Limb Health and
302 Frustration scale [3]. In the study by Legro et al., scores were higher in patients who
303 were younger than 40 years old [3]. The present study found low negative correlation
304 between age and Usefulness and Ambulation. The negative value indicates that as the
305 age of the patient increases, the worse the score gets. Locomotor activity might be
306 decreased in older individuals, which could explain the correlation. Interestingly, time
307 from prosthesis fitting to assessment had no correlation with the PEQ score, which
308 supports the findings reported by Legro and colleagues [3]. Both general health and
309 general pain correlated strongly with the scales of Usefulness, Ambulation, Transfers,
310 Perceived Responses, Social Burden and Well-being. Previous psychometric studies of
311 the PEQ have not assessed scale correlations with separate measurements of general
312 health or general pain [3-8]. However, the Usefulness scale correlated well with General
313 Health summary score in the study by Benavent et al.[8], which also supports the
314 findings of the present study.

315

316 The authors found a notable relationship between the scales of Ambulation, Social
317 burden, Usefulness and Well-being and the 15D HRQoL index in the construct validity
318 analysis [Figure 1]. The evidence suggests that PEQ has good criteria validity when it
319 comes to assessing HRQoL. Previously there has been no validated prosthesis-related
320 quality of life instrument in Finnish. Evidence of validity of the PEQ presented here
321 supports its use to assess the HRQoL of patients who have undergone major lower
322 extremity amputation and have been fitted with prosthesis. Legro et al. found strong
323 correlation with Ambulation and the SF-36 summary score of Physical Function ($r=$
324 0.61) [3]. Further, Benavent and others [8] found that there was strong correlation

325 between the Ambulation scale and the SF-36 summary scores of General Health ($r=$
326 0.71), Vitality ($r= 0.73$), Social Function ($r= 0.78$ and Mental Health ($r= 0.67$). A strong
327 correlation ($r= 0.73$) between the PEQ Social Burden scale and the SF-36 Social
328 Function summary score was also found. The authors used the 15D HRQoL instrument
329 in the present study as it is widely accepted in health care internationally and especially
330 in Finland. The 15D can be linked to the ICF-classification [23]. Its properties have
331 proven superior to several other widely used HRQoL patient-reported instruments
332 [13,14, 24,25,26].

333

334 *Clinical applications*

335 Amputation has a significant impact on patients' lives. Optimally, rehabilitation allows
336 the patients to return to their previous daily activities and social affairs. However,
337 prosthesis fitting and rehabilitation cause notable cost to society. There is a need for
338 assessment tools in measuring the need of treatment and rehabilitation as well as their
339 effectiveness. Several different techniques (e.g. in flap design) are used for major lower
340 extremity amputation. Furthermore, the amputation stump may not always have a
341 sufficient amount of healthy soft tissue for local flap stump coverage and microvascular
342 reconstruction or bone-lengthening techniques are thus needed in selected cases. These
343 surgical techniques may have an impact on how the prosthesis fits. Inadequate
344 rehabilitation methods may lead to poor results and abandonment of the prosthesis. The
345 effectiveness of different surgical methods and rehabilitation processes and their impact
346 on health-related quality of life can be assessed using the PEQ instrument in patients
347 who have been fitted with prosthesis. However, the PEQ is a comprehensive
348 questionnaire that has a large amount of items ($N = 82$). It gives extensive information

349 about the patient and prosthesis use. The 10 validated scales might be better in clinical
350 practice as they can be used as a patient profile.

351

352 ***Strengths and limitations***

353 The study recruited a heterogeneous population of patients who had undergone major
354 lower extremity amputation. Some may consider this approach as a weakness. However,
355 a heterogeneous study population allows a better generalization to be made about the
356 outcomes of this study. One limitation was the low response rate that, nonetheless, can
357 be considered acceptable for a psychometric study. Previous studies have shown that
358 ischaemia is the major cause or major lower extremity amputation [27]. However, no
359 epidemiological studies have been conducted to provide information of the amputation
360 etiology of patients who are fitted with prosthesis. Using several reference outcomes
361 would have brought even deeper knowledge of the convergent validity of the Finnish
362 PEQ. However, the authors did not have another validated amputee-specific instrument
363 in Finnish to compare. Francihignoni et al. analyzed the PEQ Ambulation scale using
364 item response theory [4]. A single item was omitted and a 5-point answer scale
365 established [4]. A Rasch analysis could have provided even more insight into the
366 construct validity of the Finnish PEQ in the present study. Further studies should
367 therefore aim to assess the construct of the PEQ scales using *inter alia* Rasch analysis
368 and the responsiveness with a longitudinal study design.

369

370 **Conclusions**

371 The authors conclude that the PEQ instrument was successfully translated and cross-
372 culturally adapted into the Finnish language version. Psychometric testing of the
373 Finnish version of the PEQ showed evidence of its reliability and validity in assessing

374 prosthesis-related quality of life in patients who have undergone major lower extremity
375 amputation and who have rehabilitated to prosthesis users. The Finnish PEQ is a
376 suitable patient-reported outcome instrument for clinical use and in scientific studies for
377 assessing the efficacy and outcomes of different amputation techniques, stump
378 reconstruction methods, and rehabilitation in patients who have been fitted with
379 prosthesis.

380

381 **Declaration of Interests** The authors report no conflicts of interest.

382

383 **References**

- 384 1. Hawkins AT, Henry AJ, Crandell DM, Nguyen LL. A systematic review of
385 functional and quality of life assessment after major lower extremity amputation. *Ann*
386 *Vasc Surg.* 2014;28(3):763-80.
- 387 2. Gauthier-Gagnon C, Grisé MC. Prosthetic profile of the amputee questionnaire:
388 validity and reliability. *Arch Phys Med Rehabil* 1994;75(12):1309-14.
- 389 3. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis
390 evaluation questionnaire for persons with lower limb amputations: assessing prosthesis-
391 related quality of life. *Arch Phys Med Rehabil* 1998;79(8):931-8.
- 392 4. Franchignoni F, Giordano A, Ferriero G, Orlandini D, Amoresano A, Perucca L.
393 Measuring mobility in people with lower limb amputation: Rasch analysis of the
394 mobility section of the prosthesis evaluation questionnaire. *J Rehabil Med.*
395 2007;39(2):138-44.
- 396 5. Day SJ, Buis A. Cross cultural equivalence testing of the Prosthetic Evaluation
397 Questionnaire (PEQ) for an Arabic speaking population. *Prosthet Orthot Int.*
398 2012;36(2):173-80.

- 399 6. Conrad C, Chamlian TR, Ogasowara MS, Pinto MA, Masiero D. Translation into
400 Brazilian Portuguese, cultural adaptation and validation of the Prosthesis Evaluation
401 Questionnaire. *Jornal Vascular Brasileiro*. 2015;14(2):110-114.
- 402 7. Safer VB, Yavuzer G, Demir SO, Yanikoglu I, Guneri FD. The prosthesis evaluation
403 questionnaire: reliability and cross-validation of the Turkish version. *J Phys Ther Sci*.
404 2015;27(6):1677-80.
- 405 8. Benavent JV, Igual C, Mora E, Antonio R, Tenias JM. Cross-cultural validation of
406 the Prosthesis Evaluation Questionnaire in vascular amputees fitted with prostheses in
407 Spain. *Prosthet Orthot Int*. 2015;40(6):713-9.
- 408 9. Boone DA, Coleman KL. Use of the Prosthesis Evaluation Questionnaire. *JPO*.
409 2006;18(6):P68-P79.
- 410 10. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A et al.
411 Principles of good practice for the translation and cultural adaptation process for
412 patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for
413 Translation and Cultural Adaptation. *Value Health* 2005;8:94-104.
- 414 11. Dewolf L, Koller M, Velikova G et al. Bottomley A. EORTC Quality of life group
415 translation procedure. 3rd ed., 2009.
- 416 12. Sintonen H. The 15D instrument of health-related quality of life: properties and
417 applications. *Ann Med* 2001;33:328–36.
- 418 13. Stavem K. Reliability, validity and responsiveness of two multiattribute utility
419 measures in patients with chronic obstructive pulmonary disease. *Qual Life Res*.
420 1999;8:45–54.
- 421 14. Alanne S, Roine RP, Räsänen P, Vainiola T, Sintonen H. Estimating the minimum
422 important change in the 15D scores. *Qual Life Res*. 2015;24(3):599-606.

- 423 15. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue
424 scales as ratio scale measures for chronic and experimental pain. *Pain* 1983;17:45–56.
- 425 16. Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, Tyrer P. Rating scales,
426 scales of measurement, issues of reliability: resolving some critical issues for clinicians
427 and researchers. *J Nerv Ment Dis* 2006;194(8):557-64.
- 428 17. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*.
429 1951;16(3):297-334.
- 430 18. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are
431 available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
- 432 19. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and
433 internal consistency. *J Pers Assess* 2003;80:99-103.
- 434 20. Streiner DL. *Health measurement scales. A practical guide to their development and*
435 *use* 3rd ed. New York: Oxford Medical Publications; 2003. 7p.
- 436 21. Beurskens AJ, de Vet HC, Koke AJ, van der Heijden GJ, Knipschild PG. Measuring
437 the functional status of patients with low back pain. Assessment of the quality of four
438 disease-specific questionnaires. *Spine*. 1995;20:1017–28.
- 439 22. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the
440 repeatability coefficient when calculating test-retest reliability. *PLoS One*.
441 2013;8(9):e73990.
- 442 23. Neljän elämänlaatumittarin sisältöanalyysi ICF-luokituksen viitekehityksessä. *Journal*
443 *of Social Medicine*. 2009;46(3):196-207.
- 444 24. Hawthorne G, Richardson J, Day NA. A comparison of the assessment of quality of
445 life (AQoL) with four other generic utility instruments. *Ann Med* 2001;33: 358–370.
- 446 25. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures:
447 results from rehabilitation patients with musculoskeletal, cardiovascular, or

- 448 psychosomatic disorders. *Qual Life Res* 2008;17:485–495.
- 449 26. Richardson J, Iezzi A, Khan MA, Chen G, Maxwell A. Measuring the Sensitivity
450 and Construct Validity of 6 Utility Instruments in 7 Disease Areas. *Med Decis Making*.
451 2016;36:147-59.
- 452 27. Eskelinen E, Lepäntalo M, Hietala EM, Sell H, Kauppila L, Mäenpää I, Pitkänen J,
453 Salminen-Peltola P, Leutola S, Eskelinen A, Kivioja A, Tukiainen E, Lukinmaa A,
454 Brasken P, Railo M. Lower limb amputations in Southern Finland in 2000 and trends up
455 to 2001. *Eur J Vasc Endovasc Surg*. 2004;27(2):193-200.
- 456
- 457

458 Table 1. Predefined hypotheses and their confirmation of rejection.

	Statistical Method	Rejected/ Confirmed
<i>Reliability</i>		
The floor and ceiling values are $\leq 15\%$	Max or min scores in %	0/10*
Internal consistency 0.80-0.90	Cronbach's alpha	4/6*
<i>Criterion validity</i>		
Moderate correlation between time of amputation and beginning of prosthesis use	Pearson	10/0*
Moderate correlation with general pain	Pearson	3/7*
Moderate correlation with general health	Pearson	4/6*
<i>Convergent validity</i>		
Large correlation between the 15D and	Standardized regression	
Ambulation	coefficients β .	Confirmed
Well-being		Confirmed

459 *Presents the number of confirmed and rejected hypotheses for all the 10 scales. β , beta.

460 Table 2. Participants' sociodemographic and clinical characteristics.

Characteristics	N = 122	461 462
Men, n (%)	76 (62.3)	463
Age, years, mean (SD; range)	63.7 (13.9;19-93)	464
Time since amputation, years, median (IQR)	4.6 (6.0)	465
Level of amputation, n (%)		466
Transtibial	81 (66.4)	467
Transfemoral	41 (33.6)	468
Bilateral amputation, n (%)	11 (9.0)	469
Indication for amputation, n (%)		470
Vascular disease	36 (29.5)	471
Trauma	25 (20.5)	472
Infection	17 (13.9)	473
Cancer	14 (11.5)	474
Other	30 (24.6)	475
Patient-reported comorbidities, n (%)		476
Diabetes	44 (36.0)	477
Vascular disease	41 (33.6)	478
Hypertension	39 (32.0)	479
Heart disease	29 (23.8)	480
Neurological disease	10 (8.2)	481
Respiratory disease	5 (4.1)	482
Other	52 (42.6)	483
General Health, VAS, mm, mean (SD)	35.71 (23.7)	
General Pain, VAS, mm, mean (SD)	34.0 (25.9)	
15D, mean score (SD)	0.820 (0.125)	

IQR, interquadrant range; SD, standard deviation;

VAS, visual analogue scale

482 Table 3. Mean scores, floor and ceiling effects and the internal consistency of each of
 483 the scales at first administration.

	Items	Response Rate (%)	Mean Score (SD)	Score Range	Floor Effect (%)	Ceiling Effect (%)	Internal Consistency (95% CI)*
<i>Prosthesis function</i>							
Usefulness	8	100	64 (19)	7-95	0	0	0.87 (0.83 to 0.92)
Residual Limb Health	6	100	60 (22)	10-98	0	0	0.79 (0.70 to 0.89)
Appearance	5	100	62 (21)	4-99	0	0	0.67 (0.52 to 0.82)
Sounds	2	98	66 (27)	5-100	0	2	0.82 (0.71 to 0.92)
<i>Mobility</i>							
Ambulation	8	100	52 (28)	0-96	1	0	0.96 (0.95 to 0.97)
Transfers	5	100	66 (25)	1-99	0	0	0.81 (0.75 to 0.88)
<i>Psychosocial experience</i>							
Perceived Responses	5	100	83 (17)	14-100	0	5	0.69 (0.55 to 0.83)
Frustration	2	96	65 (30)	2-100	0	3	0.85 (0.76 to 0.93)
Social Burden	3	98	67 (25)	3-100	0	2	0.75 (0.65 to 0.84)
<i>Well-being</i>							
Well-being	2	99	66 (23)	3-100	0	1	0.80 (0.68 to 0.91)

484 *Expresses the expected maximum size of 95% of the absolute differences between
 485 paired observations. 95% CI obtained by bias corrected and accelerated bootstrapping.
 486

487 Table 4. The change between the two measurements and reproducibility of each
 488 separate PEQ scales.

	Change From First to	Reproducibility	
	Second Measurement	ICC (95% CI)*	CR (95% CI)**
	Mean (95% CI)		
<i>Prosthesis function</i>			
Usefulness	0.6 (-1.2 to 2.3)	0.87 (0.82 to 0.91)	19 (17 to 23)
Residual Limb Health	2.1 (0.4 to 4.7)	0.80 (0.73 to 0.86)	28 (24 to 31)
Appearance	0.9 (-1.1 to 3.0)	0.85 (0.79 to 0.89)	22 (19 to 27)
Sounds	1.7 (-1.6 to 4.9)	0.80 (0.72 to 0.86)	34 (28 to 40)
<i>Mobility</i>			
Ambulation	1.9 (-0.7 to 4.5)	0.87 (0.82 to 0.91)	28 (23 to 34)
Transfers	1.5 (-1.0 to 4.0)	0.83 (0.77 to 0.88)	27 (22 to 35)
<i>Psychosocial experience</i>			
Perceived Responses	0.0 (-1.9 to 2.0)	0.78 (0.70 to 0.84)	21 (16 to 26)
Frustration	0.4 (-3.0 to 3.9)	0.81 (0.73 to 0.86)	36 (30 to 43)
Sosial Burden	2.0 (-0.9 to 4.8)	0.79 (0.71 to 0.85)	31 (26 to 35)
<i>Well-being</i>			
Well-being	0.4 (-2.6 to 3.0)	0.79 (0.71 to 0.85)	28 (24 to 32)

489 ICC, intraclass correlation coefficient; CR, coefficient of repeatability. *Obtained by
 490 one-way random-effects model with absolute agreement. **Expresses the expected
 491 maximum size of 95% of the absolute differences between paired observations. 95% CI
 492 obtained by bias corrected and accelerated bootstrapping.

493

494

495 Table 5. PEQ correlation with age, time between prosthesis and the assessment, and

496 general pain and health on visual analogue scale.

PEG Scale	Age	Time Since Amputation	General Pain	General Health
<i>Prosthesis function</i>				-
Usefulness	-0.28*	0.05	-0.39***	-0.40***
Residual Limb Health	0.23	0.00	-0.30**	-0.25
Appearance	0.15	-0.05	-0.23	-0.17
Sounds	0.24	-0.18	-0.11	-0.18
<i>Mobility</i>				
Ambulation	-0.27*	0.18	-0.44***	-0.48***
Transfers	-0.19	0.14	-0.40***	-0.40***
<i>Psychosocial experience</i>				
Perceived responses	-0.04	0.17	-0.45***	-0.42***
Frustration	0.04	0.10	-0.10	-0.05
Social Burden	-0.23	0.19	-0.40***	-0.38***
<i>Well-being</i>				
Well-being	-0.12	0.11	-0.48***	-0.43***

497 *p<0.05; **p<0.001; p<0.0001; statistical significance calculated using Sidak-adjusted

498 probabilities.

499

500 Figure 1. Predictors of the 15D age- and gender-standardized regression coefficients β .

501 Values 0.10, 0.30 and 0.50 represent small, moderate and large correlations,

502 respectively. The box plot indicates mean values and the whiskers represent standard

503 deviations.