

## Reproducibility of the Portuguese version of the PEDro Scale

### Reprodutibilidade da Escala de Qualidade PEDro em português

Silvia Regina Shiwa <sup>1</sup>  
 Leonardo Oliveira Pena Costa <sup>1,2</sup>  
 Luciola da Cunha Menezes Costa <sup>2</sup>  
 Anne Moseley <sup>2</sup>  
 Luiz Carlos Hespanhol Junior <sup>1</sup>  
 Roberta Venâncio <sup>1</sup>  
 Cintia Ruggero <sup>1</sup>  
 Tatiana de Oliveira Sato <sup>3</sup>  
 Alexandre Dias Lopes <sup>1</sup>

#### Abstract

*The objective of this study was to test the inter-rater reproducibility of the Portuguese version of the PEDro Scale. Seven physiotherapists rated the methodological quality of 50 reports of randomized controlled trials written in Portuguese indexed on the PEDro database. Each report was also rated using the English version of the PEDro Scale. Reproducibility was evaluated by comparing two separate ratings of reports written in Portuguese and comparing the Portuguese PEDro score with the English version of the scale. Kappa coefficients ranged from 0.53 to 1.00 for individual item and an intraclass correlation coefficient (ICC) of 0.82 for the total PEDro score was observed. The standard error of the measurement of the scale was 0.58. The Portuguese version of the scale was comparable with the English version, with an ICC of 0.78. The inter-rater reproducibility of the Brazilian Portuguese PEDro Scale is adequate and similar to the original English version.*

*Evidence-Based Practice; Clinical Trials; Questionnaires*

The Physiotherapy Evidence Database (PEDro; <http://www.pedro.org.au>) is a free database of randomized controlled trials (RCTs), systematic reviews and clinical practice guidelines in physiotherapy <sup>1</sup>. PEDro is considered to be one of the most comprehensive databases for indexing reports of RCTs that investigate the effects of physiotherapy interventions <sup>2</sup>. To be included in PEDro a trial must satisfy five eligibility criteria: (1) the trial must involve comparison of at least two interventions; (2) at least one of the interventions evaluated must be currently part of physiotherapy practice; (3) the interventions should be applied to subjects who are representative of those to whom the intervention might be applied in the course of physiotherapy practice; (4) the trial should involve random allocation or intended-to-be-random allocation of subjects to interventions; (5) the study must be a full paper published in a peer reviewed journal <sup>1</sup>.

All trials indexed on PEDro are assessed for methodological quality and statistical reporting using the *PEDro Scale* <sup>1</sup> that considers the following criteria: (1) eligibility criteria and source of participants; (2) random allocation; (3) concealed allocation; (4) baseline comparability; (5) blinding of subjects; (6) blinding of therapists; (7) blinding of assessors; (8) measures of key outcomes from more than 85% of participants; (9) intention-to-treat analysis; (10) between-group statistical comparisons; (11) point measures and measures of variability. The scale is

<sup>1</sup> Universidade Cidade de São Paulo, São Paulo, Brasil.

<sup>2</sup> The George Institute for Global Health, Sydney, Australia.

<sup>3</sup> Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, São Carlos, Brasil.

#### Correspondence

L. O. P. Costa  
 Universidade Cidade de São Paulo.  
 Rua Cesário Galeno 448,  
 São Paulo, SP  
 03071-000, Brazil.  
[licos3060@gmail.com](mailto:licos3060@gmail.com)

scored out of 10 with one point being awarded for each of the items (2) to (11) clearly satisfied and reported by the trial. Each trial is assessed by two independent raters. Any disagreements are arbitrated by a third rater. Reports of trials written in languages other than English are assessed by bilingual raters using the English version of the *PEDro Scale*.

More than 4,300 searches of PEDro are performed everyday by physiotherapists worldwide, of which 15% are performed by professionals from Portuguese-speaking countries. Although the PEDro website was recently translated into Portuguese, the search and results pages, including the *PEDro Scale* score, are only available in English. As only a small proportion of the Brazilian population are fully proficient in English, it is likely that this limitation of the PEDro search function will restrict the use of PEDro by Portuguese-speaking physiotherapists.

As a first step to making PEDro more accessible to physiotherapists that are not proficient in English, the *PEDro Scale* has been cross-culturally adapted into Portuguese by a Portuguese research group<sup>3</sup>. This adaptation was performed following the recommendations from current guidelines<sup>4</sup>.

Some orthographical adjustments were made to the Portuguese version of the *PEDro Scale* to adapt it to Brazilian Portuguese. No semantic or grammatical modifications were necessary. The scale in Brazilian Portuguese can be accessed at: <http://www.pedro.org.au/portuguese/downloads/pedro-scale/>. To date, the reproducibility of the *Brazilian Portuguese PEDro Scale* has not been evaluated.

The objective of this study was to test the reproducibility and parallel-form reproducibility of the Brazilian Portuguese version of the *PEDro Scale* compared to the original English version.

## Methods

Seven Brazilian physiotherapists participated in this study as independent raters. The raters underwent training to assess the methodological quality of RCTs using the Brazilian Portuguese version of the *PEDro Scale* through the online PEDro Scale Training Program (<http://training.pedro.org.au/portuguese/index.html> – this website is password protected). After training, each therapist had to pass an accuracy test to ensure that they were applying the scale to the same standard as other raters.

Fifty reports of RCTs written in Portuguese were used in this study derived from a universe of 65 RCTs indexed on the database. The cita-

tion details and PEDro ratings for all Portuguese language trials indexed on PEDro were downloaded on 2 August 2010. Two sets of consensus ratings (individual items and total PEDro score) were generated by the Brazilian raters using the Brazilian Portuguese version of the *PEDro Scale* (Figure 1). To produce the first set, each trial was independently assessed by two raters and any disagreements were arbitrated by a third rater. This process was then repeated by three different raters to produce a second set of consensus ratings.

The reproducibility of the Brazilian Portuguese version was evaluated based on the definition from the guidelines for health status questionnaires: “the degree to which repeated measurements provide similar answers in stable conditions”<sup>5</sup> (p. 36). Both reliability (“the degree to which evaluations could be distinguished from each other, despite measurement error”) and agreement (“how close repeated measures are, expressed in the unit of the scale being tested”) were investigated. The two sets of the Brazilian-Portuguese version consensus ratings were compared to evaluate test-retest reproducibility. The first set of Brazilian Portuguese consensus ratings was compared to the English version consensus ratings to evaluate parallel-form reproducibility. A sensitivity analysis was also performed to compare the second set of Brazilian Portuguese consensus ratings with those of the English *PEDro Scale*.

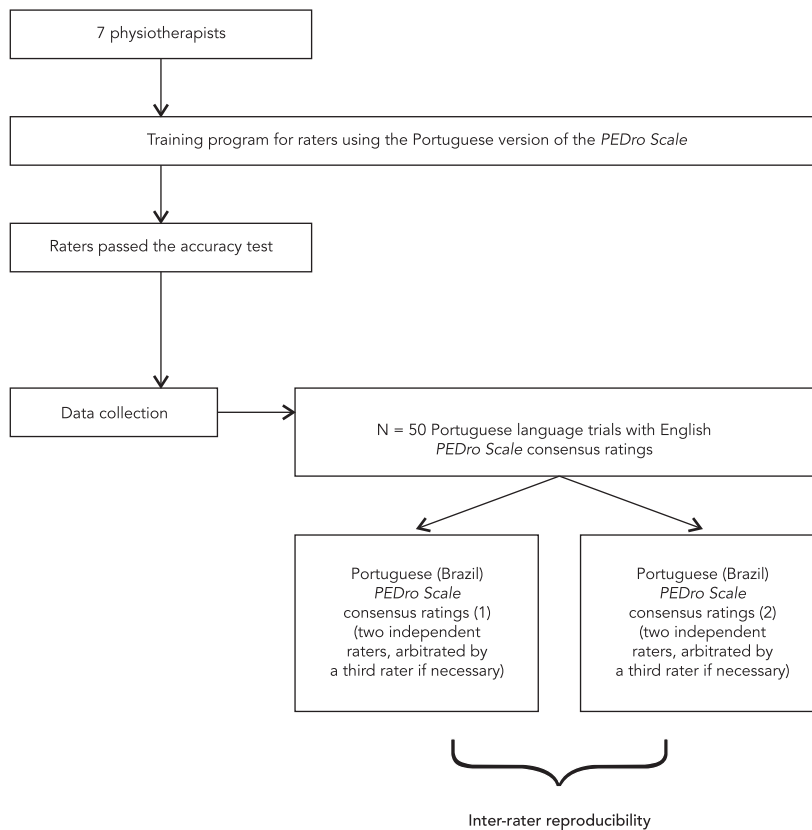
The following statistics were determined to evaluate the test-retest reproducibility of the Brazilian-Portuguese version of the *PEDro Scale*: 1) kappa coefficients for each item of the *PEDro Scale* (kappa estimates for an individual item are considered important so that PEDro users can critically appraise the reproducibility of each item from the *PEDro Scale*); 2) intraclass correlation coefficient type 1,1 ( $ICC_{1,1}$ )<sup>6</sup> and its 95% confidence intervals (95%CI) for the total PEDro score and to determine parallel-form reproducibility; 3) standard error of measurement (SEM) was calculated by dividing the standard deviation of the difference in the total PEDro scores by the square root of two<sup>5</sup>.

## Results

The mean total PEDro score for the 50 reports of trials was 3.5 (SD = 1.4; range 1-7), indicating that the majority of the reports were of low methodological quality. Five of the *PEDro Scale* items were clearly satisfied in only 10% (or less) of the reports (Table 1). No *PEDro Scale* item was clearly satisfied in 90% (or more) of the reports.

Figure 1

Flow chart of the study process.



“Moderate” to “almost perfect”<sup>7</sup> test-retest reproducibility was obtained for individual *PEDro Scale* items assessed using the Brazilian Portuguese version of the *PEDro Scale* (Table 1). Kappa values of items 2, 5 and 6 were greater than 0.80 and these criteria were classified as having “almost perfect”<sup>6</sup> reliability. Kappa values of items 3, 7, 9, 10 and 11 ranged between 0.61 and 0.80 and these criteria were classified as having “substantial” reliability. Items 1, 4 and 8 were classified as having “moderate” reliability (kappa values between 0.41 and 0.60). In contrast, in the test-retest reliability of the English version of the *PEDro Scale*, three items were classified as having “moderate” reliability and eight items as having “substantial” reliability<sup>1</sup>.

The test-retest reproducibility of the total PEDro score generated using the Brazilian Portuguese version was classified as “excellent”, with an ICC<sub>1,1</sub> of 0.82 (95%CI: 0.70-0.89). This is great-

er than the value of 0.68 (95%CI: 0.57-0.76) contained for the English version of the *PEDro Scale*<sup>1</sup>. The SEM was 0.58 for the total PEDro score. This was classified as “good” agreement.

The parallel-form reproducibility between the Portuguese and English versions of the *PEDro Scale* was classified as “excellent”. The ICC<sub>1,1</sub> of the English version compared to the first Brazilian Portuguese set was 0.78 (95%CI: 0.67-0.86). The sensitivity analysis using the second set of Portuguese ratings yielded similar results with an ICC<sub>1,1</sub> of 0.75 (95%CI: 0.59-0.85).

## Discussion

It was observed that the Brazilian Portuguese *PEDro Scale* has good reproducibility. Parallel-form reproducibility between the Portuguese and English versions of the *PEDro Scale* is excellent.

Table 1

Kappa coefficients and the proportion of reports with a yes response (base rate) for each of the individual item for ratings generated using the Portuguese (Brazil) and English versions of the *PEDro Scale*.

<i>PEDro Scale</i> items		Portuguese Kappa (95%CI)	English * Kappa	Portuguese 1 (%)	Base-rate Portuguese 2 (%)	English (%)
1.	Eligibility criteria and source specified	0.57 (0.32-0.82)	0.63	80	72	72
2.	Random allocation	0.91 (0.73-1.00) **	0.79	86	88	84
3.	Concealed allocation	0.73 (0.38-1.00) **	0.70	8	8	12
4.	Baseline comparability	0.60 (0.38-0.82)	0.50	50	66	42
5.	Blinding of subjects	1.00 (1.00-1.00)	0.70	4	4	4
6.	Blinding of therapists	1.00 (1.00-1.00)	0.79	0	0	0
7.	Blinding of assessors	0.78 (0.49-1.00) **	0.79	10	10	14
8.	More than 85% follow-up	0.53 (0.29-0.77)	0.67	36	38	36
9.	Intention-to-treat analysis	0.66 (0.03-1.00) **	0.57	2	4	0
10.	Reporting of between-groups statistical comparisons	0.66 (0.44-0.88)	0.68	64	60	68
11.	Reporting of point measures and measures of variability	0.74 (0.52-0.96)	0.54	78	72	72

\* The original study did not provide 95% confidence intervals;

\*\* 95% confidence intervals (95%CI) are asymmetric because the upper-bound estimate was outside the range of the scale. In such cases the value was adjusted to 1.00.

As only orthographical differences between the Portuguese (Portugal) version of the *PEDro Scale* were made to create the Brazilian Portuguese version, the reproducibility estimates obtained can be applied to both Portuguese versions of the *PEDro Scale*.

The reliability estimates of individual items of the Brazilian Portuguese version of the *PEDro Scale* were found to be similar to the estimates obtained by a previous study that tested the reliability of the English version of the scale <sup>1</sup>. Although it is interesting to note that kappa estimates were slightly higher for the Brazilian scale (Table 1), the factors involved in this comparison are complex and any such conclusion should be treated with caution because different RCTs were evaluated.

The reliability estimates of the total *PEDro* score of the English *PEDro Scale* (ICC = 0.68) <sup>1</sup> and Portuguese version (0.82) were similar. The SEM of the Portuguese version of the *PEDro Scale* reflects a low absolute measurement error. The “excellent” parallel-form of reproducibility between the Portuguese and English versions of the *PEDro Scale* (ICC = 0.78) suggest that either version may be used to generate *PEDro* ratings.

An important incidental finding of this study is the low methodological quality and the poor statistical reporting found in the reports that were assessed. The mean total *PEDro* score of re-

ports in Portuguese was lower than that recorded by a study describing the quality of 3,120 reports of RCTs relevant to physiotherapy <sup>8</sup>. Perhaps a review of journal editorial policies, together with training provision for authors and reviewers, is required to bridge this gap in reports of RCTs written in Portuguese.

In conclusion, the scores obtained using the Brazilian Portuguese version of the *PEDro Scale* are similar to those achieved using the original English version. It is now possible for Portuguese speakers that are not proficient in English to use the *PEDro Scale* to critically appraise the methodological quality and statistical reporting of reports of RCTs.

## Resumo

*O objetivo foi testar a reprodutibilidade da versão em português da Escala de Qualidade PEDro. Sete fisioterapeutas avaliaram a qualidade metodológica de 50 estudos controlados aleatorizados em português, indexados na base de dados PEDro. Cada artigo já possuía sua respectiva avaliação nessa base de dados, utilizando a versão em inglês da escala PEDro. Foi calculada a confiabilidade da escala, assim como foi comparada a pontuação total de consenso com a pontuação das avaliações utilizando a escala em inglês. Os coeficientes kappa variaram entre 0,53 e 1,00 para itens individuais, e um coeficiente de correlação intraclassa (CCI) de 0,82 foi obtido para a pontuação total. O erro-padrão de medida foi de 0,58 ponto. A versão em português da escala foi comparada com a versão em inglês e foi observado um CCI de 0,78. A reprodutibilidade da versão em língua portuguesa da Escala de Qualidade PEDro foi adequada e similar à versão em inglês.*

*Prática Clínica Baseada em Evidências; Ensaio Clínico; Questionários*

## Contributors

L. O. P. Costa participated in the definition and management of data collection, statistical analysis and in drafting of this article. S. R. Shiwa, A. D. Lopes, L. C. M. Costa, A. Moseley, L. C. Hespanhol Junior e T. O. Sato collaborated in defining the study, data collection and drafting of this article. R. Venâncio e C. Ruggero contributed to defining the study, data collection and final revision of the manuscript.

## Acknowledgments

We are grateful for the financial support provided for this study by FAPESP.

## References

1. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther* 2003; 83:713-21.
2. Michaleff ZA, Costa LOP, Moseley AM, Maher CG, Elkins MR, Herbert RD, et al. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. *Phys Ther* 2011; 91:190-7.
3. Costa C. Portuguese adaptation of the PEDro scale. Lisboa: Universidade Técnica de Lisboa; 2011.
4. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25:3186-91.
5. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60:34-42.
6. Krebs DE. Declare your ICC type. *Phys Ther* 1986; 66:1431.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.
8. Costa LO, Moseley AM, Sherrington C, Maher CG, Herbert RD, Elkins MR. Core journals that publish clinical trials of physical therapy interventions. *Phys Ther* 2010; 90:1631-40.

Submitted on 17/Feb/2011

Final version resubmitted on 28/Jun/2011

Approved on 06/Jul/2011