






Interexaminer and test-retest reliability of the Participation Scale (P-Scale) in stroke patients

Confiabilidade interavaliadores e teste-reteste da Escala de Participação (P-Scale) em pacientes após acidente vascular cerebral

 Fabiana Caetano Martins Silva e Dutra¹,  Alessandra Carvalho Alves²,  Alberto Luiz Aramaki²,  Máira Ferreira do Amaral¹,  Alessandra Cavalcanti¹

ABSTRACT

The Brazilian version of the Participation Scale (P-Scale) is an evaluation instrument that assesses the restrictions in participation in individuals with several different health conditions. **Objectives:** Evaluating the interexaminer and test-retest reliability of the Brazilian version of the P-Scale in a group of stroke patients. **Methods:** Methodological study that evaluated 20 stroke patients attended in a Rehabilitation Center. The reliability of the total score was evaluated using the Intraclass Correlation Coefficient (ICC). The interexaminer and test-retest reliability of each item were measured using the quadratic weighted Kappa Coefficient. To evaluate the agreement between the scores obtained per each interexaminer and test-retest item, the Bland-Altman graphic was used. **Results:** The Brazilian version of the P-Scale presented an almost perfect reliability in most items ($kw > 0.81$), with no systemic measuring errors according to the Bland-Altman agreement graphic ($p = 0.350$) and an excellent total score test-retest reliability ($ICC = 0.96$; $P = 0.000$). Regarding interexaminer reliability, the items varied from almost perfect ($kw > 0.81$) to regular ($0.21 < kw < 0.40$) and the Bland-Altman agreement graphic only indicated systemic measuring differences ($p = 0.018$) in two atypical cases. The ICC of interexaminer reliability total score was excellent in both the analysis of the entire sample ($ICC = 0.95$ for 20 individuals; $p = 0.000$) and in the sample without including the atypical cases ($ICC = 0.97$ for 18 individuals; $p = 0.000$). **Conclusion:** The Brazilian version of the P-Scale presented excellent interexaminer and test-retest reliability to evaluate participation in stroke patients.

Keywords: Stroke, Social Participation, Reproducibility of Results, Validation Study

RESUMO

A versão brasileira da Escala de Participação (P-Scale) é um instrumento de avaliação que mensura restrições em participação em indivíduos com diversas condições de saúde. **Objetivos:** Avaliar a confiabilidade interexaminador e teste-reteste da versão brasileira da P-Scale em um grupo de pacientes pós-acidente vascular cerebral (AVC). **Método:** Estudo metodológico, avaliou 20 pacientes com diagnóstico de AVC atendidos em um Centro de Reabilitação. A confiabilidade do escore total foi avaliado por meio do Coeficiente de Correlação Intraclassa (ICC). A confiabilidade interavaliador e teste-reteste de cada item foi medida pelo Coeficiente Kappa com ponderação quadrática. Para avaliação da concordância entre os escores obtidos item por item interavaliador e teste-reteste, foi utilizado o gráfico Bland-Altman. **Resultados:** A versão brasileira da P-Scale apresentou confiabilidade quase perfeita da maioria dos itens ($kw > 0,81$), sem erros sistemáticos de mensuração avaliados pelo gráfico de concordância de Bland-Altman ($p = 0,350$) e excelente confiabilidade teste-reteste do escore total ($ICC = 0,96$; $p = 0,000$). Em relação à confiabilidade interexaminador, os itens apresentaram confiabilidade variando de quase perfeita ($kw > 0,81$) a regular ($0,21 < kw < 0,40$) e o gráfico de concordância de Bland-Altman indicou diferenças sistemáticas de mensuração ($p = 0,018$) apenas para dois casos atípicos. ICC da confiabilidade interexaminador do escore total foi excelente tanto na análise da amostra total ($ICC = 0,95$ para 20 indivíduos; $p = 0,000$), quanto na amostra sem os casos atípicos ($ICC = 0,97$ para 18 casos; $p = 0,000$). **Conclusão:** A versão brasileira da P-Scale apresentou excelente confiabilidade teste-reteste e interexaminador para avaliar participação em pacientes após AVC.

Palavras-chaves: Acidente Vascular Cerebral, Participação Social, Reprodutibilidade dos Testes, Estudo de Validação

¹Universidade Federal do Triângulo Mineiro - UFTM

²Núcleo de Estudos e Pesquisas em Trabalho, Participação Social e Saúde - NETRAS

Address for correspondence

Fabiana Caetano Martins Silva e Dutra
E-mail: fabiana.dutra@uftm.edu.br

Submitted: June 4, 2020

Accepted: December 30, 2021

How to cite

Dutra FCMS, Alves AC, Aramaki AL, Amaral MF, Cavalcanti A. Interexaminer and test-retest reliability of the Participation Scale (P-Scale) in stroke patients. *Acta Fisiátrica*. 2022;29(1):42-49.



10.11606/issn.2317-0190.v29i1a168523



©2022 by Acta Fisiátrica

This work is licensed under a Creative Commons - Attribution 4.0 International

INTRODUCTION

Participation refers to a state of health and functionality where a person can get involved in real-life roles and situations.¹ Participation includes personal activities, mobility in different environments, social relationships, work, education, leisure, spirituality, community, and citizen life.¹ This concept comprises the experiences of the individual, his involvement, and the contextual, personal, and environmental factors surrounding the individual.² Therefore, participating means taking part, being included, or getting involved in an area of life, as well as being accepted or having access to the necessary resources for this inclusion.^{1,2}

According to the World Health Organization (WHO), participation is influenced by demographical factors (such as age, sex, education) and environmental aspects (physical, social, and political).¹ The interactions among health, personal, and environmental factors impact the level of involvement in daily activities and socially expected roles, influencing participation.² Restrictions on participation are a common problem among people with disabilities and those undergoing rehabilitation treatment, such as patients with stroke sequelae.³⁻⁵

Stroke is the leading cause of acquired physical disability in adults worldwide and represents a critical public health issue considering its prevalence, severity, and cost. After a stroke episode, about 65% of patients report restrictions on reintegration into community activities, and 54% have high limitations to perform activities of daily living, household chores, and shopping,⁶ a characteristic that endures until the chronic phase of stroke. After four years, almost half of stroke survivors have some activity limitation, and about 30% of them are classified as restricted in participation.⁷ Evidence shows that persistent restrictions on participation have a significant impact on quality of life, the degree of dependence to perform activities of daily living, on social interaction, and the increase of disabilities in people with stroke sequelae.^{5,8,9}

Measuring the degree of restriction in post-stroke participation is meaningful to determine the impact of disability on these individuals, provide assistance to health professionals when planning their interventions, and measure the effectiveness of interventions designed to reduce the consequences of disability. Therefore, restrictions on participation are relevant information that must be considered throughout the rehabilitation process in post-stroke people. The documentation of this construct through assessments that identify the functional needs of patients is applicable for both clinical interest and scientific research.

In developing countries, information on the participation of individuals with rehabilitation needs is limited. In this context, the International Classification of Functioning, Disability, and Health (ICF) of the World Health Organization (WHO) provides health professionals with a conceptual model that helps clinical reasoning and the development of assessment instruments from a biopsychosocial perspective of health.¹⁰ Due to the broad approach and application of the ICF, new assessment scales have been developed to assess and measure the domains and components of functionality. A recent example is the Participation Scale (P-Scale), which evaluates individual participation in different health conditions and cultural

environments.¹¹

In Brazil, studies have applied the P-Scale to a large population with disabilities treated with different rehabilitation services.^{2,4,12-16} A national study tested the usefulness of the P-Scale in rehabilitation services. It provided helpful information for improving this assessment, demonstrating the most difficult items for patients to respond to and proposing an amplification of the scale by adding items that fit the restriction of these individuals.¹⁴ In the worldwide literature, different studies show good psychometric properties and validity of the P-Scale.¹⁷⁻¹⁹ However, no study has evaluated its reliability in Brazil, specifically for post-stroke patients.

OBJECTIVE

This study aims to test the inter-rater reliability with a test-retest method of the Brazilian version of the P-Scale in a sample of patients with stroke sequelae at a public rehabilitation service.

METHOD

This is a methodological study for establishing the inter-examiner reliability and test-retest of the P-Scale, approved by the Research Ethics Committee of the Federal University of Triângulo Mineiro (UFTM) (CAAE: 46357215.2.0000.5154).

Following Hobart et al.²⁰ and considering a precision of 0.90, an estimated error of 0.2, and an alpha of 0.05, 20 patients diagnosed with stroke treated at a public rehabilitation service were randomly selected.

Patients over 18 years of age, with stroke diagnosis for more than six months and less than five years, without other neurological or orthopedic comorbidities, capable of understanding and answering the questionnaires with reliability were eligible for this study.

Participation Scale (P-Scale) description

The Participation Scale (P-Scale) is an 18-item assessment tool that enables the quantification of self-reported restrictions on the participation of individuals with some health condition or disability comprised in eight of the nine main areas of life defined by the ICF (Learning and applying knowledge; Communication and Self-care; Mobility; Domestic life; Interpersonal interactions; Community, social and civic life).¹¹ During the application, the interviewee must compare himself with an actual or hypothetical 'peer', someone similar in all aspects, except for the disability.¹ They are also asked if their level of participation is equal to or less than that of their hypothetical peers. Then, if the respondent reports a possible difficulty, they are requested to indicate the degree to which this difficulty poses a problem in their daily life.

The score for each item lies on a 5-point Likert scale where 0 (zero) means the interviewee does not consider his participation to be lesser than his "peer", 1 means their participation is low, but this is not a problem, 2 means this is a small problem, 3 means a medium problem, and 5 means a big problem. The values of each item are added and the total value of the P-Scale score varies between 0 (zero) and 90. Lower values indicate that the respondent has fewer restrictions on

their participation.¹¹ The value assumed as the maximum limit for the individual to be considered free of restrictions on participation is 12 points, and higher scores indicate some degree of restriction on participation.¹¹ Patients with values between 13 and 22 points have mild restriction, values between 23 and 32 suggest moderate restriction, values between 33 and 52 imply patients with severe restriction, and values above 53 points indicate extreme restriction for participation.¹¹

Procedures

Two raters were trained in using the P-Scale by an occupational therapist with experience in using the instrument for applying the questionnaire. The training consisted of reading the Portuguese version of the P-Scale manual and discussing how to apply it. None of the evaluators had experience using the assessment or previous access to the manual. The data collection was carried out in a public rehabilitation service from February to July 2017.

The data collection procedure was conducted in two stages. In the first, sociodemographic and clinical information such as gender, age, risk factors, self-perception of health, employment status, family status, prostheses and/or orthosis, and dependence to perform daily activities were recorded. After that, the participants answered the P-Scale as an interview by examiner A. This first stage lasted about 30 minutes. Then, to analyze inter-rater reliability, the participant rested for 30 minutes and examiner B evaluated the participants by reapplying the P-Scale, spending an average time of 20 minutes.

Seven to ten days after the first assessment, the second stage of the procedures started and the participants were reassessed by examiner A. The estimated time of this step was 20 minutes and allowed the P-Scale test-retest reliability. The examiners did not have access to the results of the first step assessments, so there would be no memorization of the results.

Data analysis

Descriptive statistics characterized the sample. The partial reliability of each question and the total P-Scale score were addressed in the test-retest and inter-rater reliability analysis. The test-retest and inter-rater reliability of each item was investigated with the Kappa Coefficient with quadratic weighting (kw). The interpretation of the Kappa agreement level followed the criteria proposed by Landis & Koch,²¹ as follows: almost perfect, 0.81 to 1.00; strong: 0.61 to 0.80; moderate: 0.41 to 0.60; regular: 0.21 to 0.40; discrete: 0 to 0.20; and poor: < 0.

A Bland-Altman plot was generated, calculating the mean difference between the two measures (test-retest or inter-rater) and the 95% confidence interval to determine the degree of agreement between the scores, the disagreement magnitude, and the existence of errors and systematic patterns between both measures of each item in the test-retest and inter-rater analysis.²²

The degree of reliability of the P-Scale total score was established with the Intraclass Correlation Coefficient (ICC), whose interpretation followed the criteria proposed by

Munro:²³ very low, 0 to 0.25; low, 0.26 to 0.49; moderate, 0.50 to 0.69; high 0.70 to 0.89; and very high, 0.90 to 1.00.

All analyzes were conducted with the SPSS IBM® software (Statistical Package for the Social Science), version 20.0.

RESULTS

Twenty patients were included and evaluated. They were predominantly males (55%) with a mean age of 48.20 years (SD= 15.68), mostly married (60%), living with a spouse and children (65%), and 70% were not responsible for the family income. Most of the participants had up to eight years of schooling (65%), and 25% had completed elementary school. Most participants were on a work leave (55%) being assisted by the Brazilian National Institute of Social Security, and six (30%) were retired.

Regarding life habits, 65% had some risk factors such as being overweight, a sedentary lifestyle, alcohol consumption, or smoking, whereas 50% considered their physical health as good or very good, and 65% reported their emotional health as good or very good. Most participants (75%) needed assistive devices, especially glasses or contact lenses 35%, and 60% needed some level of help to carry out daily activities, among which 65% were assisted by their children, spouses, or parents.

All respondents receive some treatment or medication. Table 1 presents the complete description of the social and health status of the participants in this study.

Test-retest reliability

Regarding the total P-Scale score, in the first evaluation, examiner A obtained a mean of 33.45 points (SD= 20.74), ranging from 6 to 82 points. Categorizing the questionnaire answers, 4 participants (20%) had no restrictions, 2 (10%) had mild restrictions, 4 (20%) had moderate restrictions, 7 (35%) had severe restrictions, and only 3 (15%) reported extreme restrictions. The second P-Scale application generated scores from 2 to 80 points and a mean of 30.85 (SD= 21.32). The classification showed 6 participants (30%) with no restrictions, 3 (15%) with mild restrictions, 1 (5%) with moderate restrictions, 7 (35%) with severe restrictions, and 3 (15%) with extreme restrictions. The ICC analysis of the P-Scale total score agreement of the test-retest comparisons revealed a very high correlation between both evaluations (ICC= 0.96; 95%CI 0.89 - 0.98; $p < 0.001$).

In the individualized test-retest reliability analysis for each of the 18 P-Scale items, the questions 3, 4, 5, 10, 11, 13, 14, 17, and 18 achieved an almost perfect reliability ($kw > 0.81$), questions 6, 9, 12, and 17 showed strong reliability ($0.61 < kw < 0.80$), questions 1 and 8 had moderate reliability ($0.41 < kw < 0.60$), and regular reliability was found in items 2, 15, and 16 ($0.21 < kw < 0.40$). This item-by-item evaluation with Kappa coefficient and quadratic weighting is described in Chart 1.

The graphical analysis of the agreement between both evaluation moments (test-retest) indicated a mean difference of 1.94 (SD= 8.59) with a 95% CI ranging from -2.32 to 6.21. The difference between the means approaches zero, meaning the P-Scale scores found in both evaluation moments were similar, with no systematic significant measurement errors ($p = 0.350$).

The Bland-Altman agreement plot for the test-retest

moment is shown in Figure 1. As demonstrated, no participant showed a difference in the means between the two applications above CI95% of the limits of agreement, represented by the dashed lines.

Chart 1. Kappa coefficients classification for test-retest reliability analysis of P-Scale (n= 20)

Reference values for Kappa reliability coefficient	Items
>0.81 (almost perfect)	3, 4, 5, 10, 11, 13, 14, 17, 18
0.61-0.80 (strong)	6, 7, 9, 12
0.41-0.60 (moderate)	1, 8
0.21-0.40 (regular)	2, 15, 16
0.00-0.20 (discrete)	-
<0 (poor)	-

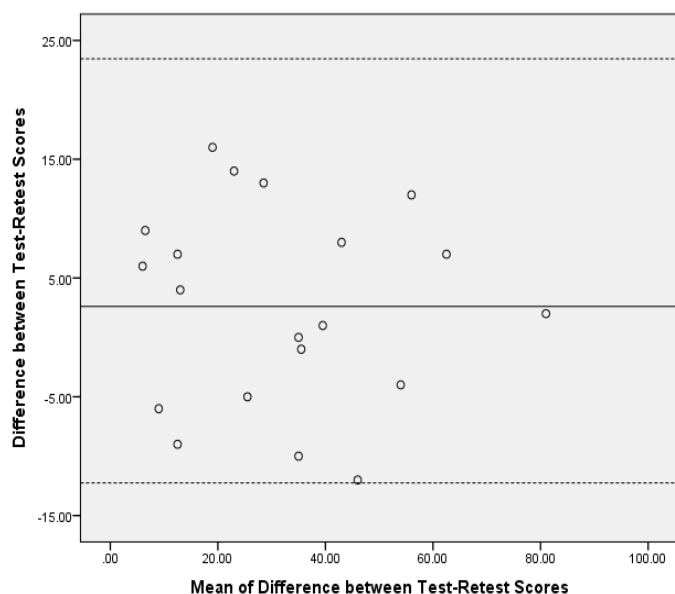


Figure 1. Bland-Altman plot for the P-Scale total score test-retest agreement (n= 20)

Inter-rater reliability

The Kappa coefficient with the quadratic weighting of the inter-rater reliability assessment for each item is reported in Chart 2. In summary, items 3, 7, 10, and 15 have almost perfect reliability ($k_w > 0.81$), items 4, 12, and 18 had strong reliability ($0.61 < k_w < 0.80$), items 1, 8, and 14 had moderate reliability ($0.41 < k_w < 0.60$), items 2, 5, 6, 9, 13, and 16 showed regular reliability ($0.21 < k_w < 0.40$), item 17 had discrete reliability ($0.00 < k_w < 0.20$), and item 11 had poor reliability ($k_w < 0.00$).

The graphical analysis of the agreement between the evaluators demonstrated the mean of the differences as -5.88

(SD= 9.58), with a 95% CI of -10.65 to -1.12, indicating that the scores found by the evaluators may present systematic measurement differences ($p = 0.018$), as the difference in means is far from zero. The Bland-Altman concordance plot for the inter-rater assessment (Figure 2) indicates the participants had significant score differences, depending on the examiners.

Therefore, the ICC test for the total P-Scale score was conducted with and without including atypical cases ($n = 20$ and $n = 18$, respectively) to address the disagreement issue. Hence, the results of the P-Scale total score measured by examiner A (mean= 33.45 points; SD= 20.74, ranging from 6 to 82 points) were compared with the results evaluated by the examiner B (mean= 39.05 points; SD= 22.31, ranging from 5 to 85 points).

In this comparison of inter-rater reliability, the ICC revealed a very high correlation in both analyses, with ICC= 0.95 with the atypical cases or outliers ($n = 20$; CI= 0.88-0.98; $p = 0.000$), and ICC= 0.97 without the outliers ($n = 18$; CI= 0.4-0.99; $p = 0.000$).

Chart 2. Kappa coefficients classification for inter-rater reliability analysis of P-Scale (n=20)

Reference values for Kappa reliability coefficient	Items
>0.81 (almost perfect)	3, 7, 10, 15
0.61-0.80 (strong)	4, 12, 18
0.41-0.60 (moderate)	1, 8, 14
0.21-0.40 (regular)	2, 5, 6, 9, 13, 16
0.00-0.20 (discrete)	17
<0 (poor)	11

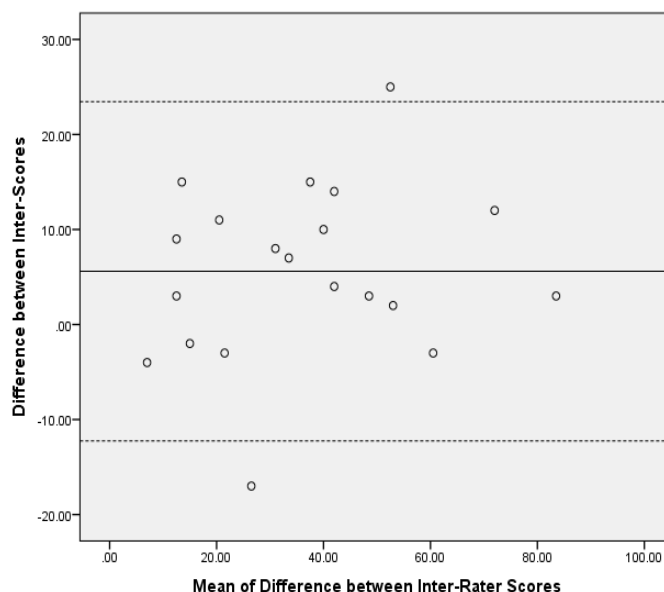


Figure 2. Bland-Altman plot for the P-Scale total score inter-rater agreement (n= 20)

Table 1. Sociodemographic and health information (n= 20)

Variable	Frequency	(%)
Sex		
Female	9	45.00%
Male	11	55.00%
Marital status		
Single	7	35.00%
Married	12	60.00%
Common-law marriage	1	5.00%
Shares household with		
None	1	5.00%
Spouse	2	10.00%
Children	1	5.00%
Parents	2	10.00%
Other relatives	1	5.00%
Spouse and children	13	65.00%
Occupation		
Employed	1	5.00%
Unemployed	2	10.00%
On leave	11	55.00%
Retired	6	30.00%
Schooling		
Illiterate	2	10.00%
Complete elementary level I	5	25.00%
Incomplete elementary level I	2	10.00%
Complete elementary level II	3	15.00%
Incomplete elementary level II	1	5.00%
Complete high school	1	5.00%
Incomplete high school	5	25.00%
Incomplete College	1	5.00%
Responsible for Family income		
Yes	6	30.00%
No	14	70.00%
Risk factors		
Smoking	2	10.00%
Sedentarism	6	30.00%
Smoking and alcohol consumption	3	15.00%
Smoking and Sedentarism	1	5.00%
Alcohol consumption and overweight	1	5.00%
No risk factor	7	35.00%
Self-evaluation of physical health		
Very good	1	5.00%
Good	9	45.00%
Moderate	3	15.00%
Bad	5	25.00%
Very bad	2	10.00%
Self-evaluation of emotional health		
Good	6	30.00%
Moderate	7	35.00%
Bad	5	25.00%
Very bad	2	10.00%
Using assistive device		
Yes	15	75.00%
No	5	25.00%
Assistive device usage		
None	5	25.00%
Glasses or contact lenses	7	35.00%
Wheelchair or shower chair	3	15.00%
Walker / Crutches / Cane	1	5.00%
Orthosis	1	5.00%
Glasses or Contact lenses and hearing aids	1	5.00%
Glasses or Contact lenses and Wheelchair or shower chair	2	10.00%
Demand assistance for ADL		
Yes	12	60.00%
No	8	40.00%
Who assists patients in ADL		
None	7	35.00%
Spouse	5	25.00%
Children	3	15.00%
Parents	5	25.00%
Ongoing treatments		
Rehabilitation professional	14	70.00%
Pharmacologic	1	5.00%
Chemotherapy or radiotherapy	1	5.00%
Medical treatment for rehabilitation issues	3	15.00%
Medical treatment for other issues	1	5.00%

DISCUSSION

Our study evaluated the reliability of the Brazilian version of the P-Scale in a group of patients with stroke sequelae treated at a public rehabilitation service. The results showed that the Brazilian version of the P-Scale has adequate test-retest and inter-rater reliability, with agreement values within acceptable limits.

Most participants were male, on work leave or retired, and with low family income. These characteristics are also found in other studies that assessed sociodemographic information of post-stroke patients.²⁴⁻²⁶ Although the occurrence of stroke is associated with aging, data from 119 countries show that the number of cases in the 20-64 age group increased from 25% in 1990 to 31% in 2010.²⁴ This trend was found among our research participants, who were generally adults with stroke onset before 60 years of age.

Most interviewees reported lifestyle habits that may be considered risk factors for developing chronic-degenerative diseases, mainly physical inactivity, a significant factor strongly correlated with the occurrence of stroke.^{24,25,27} These traits represent the characteristics of the Brazilian population, where only about 25% are physically active people, about 13% of men and 9.3% of women are smokers, and 14.9% of the adult population ingest alcoholic drinks above the limits recommended by the WHO.²⁷ Although these life habits follow the trend of national data, these results are worrying and suggest it is difficult for patients to change habits considered risk factors for chronic degenerative diseases. This information is noteworthy concerning actions to promote health in general or healthy habits by health services, especially interventions designed to prevent diseases of the vulnerable population.

The results found by the examiners in both evaluation moments showed that the average score ranged from 30 to 39 points, approximately. Regarding the degree of restriction, the values between the examiners and between the two evaluation moments varied from moderate to extreme restriction when evaluated together. Other studies conducted with a population of people with disabilities found similar values.^{2,13,16}

The ICC is used to assess the consistency of scores obtained by repeated applications, considering the sample variation.²⁸ In this perspective, the P-Scale showed good consistency for the sample studied, both in test-retest and inter-rater analyses. The inter-rater reliability showed adequate values for the total score, with a correlation of $\alpha = 0.95$, even considering atypical cases. This value was similar to those reported by Stevelink et al.¹⁸ who found indices of inter-rater reliability for the P-Scale of 0.90, and higher than the results found by Van Brakel et al.¹¹ and Van Der Zee et al.¹⁷ whose values were 0.80 and 0.82, respectively. Our results for inter-rater reliability indicate that the strategies used for the P-Scale application were adequate to maintain the consistency of the measures, regardless of the evaluator.

Likewise, the test-retest reliability results showed $\alpha = 0.96$, a considerable value of measurement consistency, compared to the original study during the development of the P-Scale.¹¹ This result suggests that examiner A was especially consistent and that the P-Scale items are sufficiently clear, contributing to the measurement suitability, without being influenced by the application moment.

The Bland and Altman plot test showed excellent test-retest agreement, indicating that the scores obtained in the first evaluation agreed with those in the second in 95% of the cases. Excellent inter-rater agreement was also observed, confirming the score agreement of examiners A and B in 95% of the items. This plot shows that no systematic error was observed, i.e., there was no tendency to underestimate or overestimate the restrictions on participation during the application of the P-Scale.

However, even with these consistent values, there were two cases of outliers observed in the scatter diagram, evidenced by two observations at the extremities of the limits of the agreement interval. Two patients presented atypical behavior with different scores between the examiners A and B. Therefore, we recommend caution in interpreting P-Scale when there is high individual variability, influencing the mean values obtained during the assessments.²⁸ Inter-rater reliability analysis of the total score was performed with and without these outliers to adjust for this situation. In both analyses, the ICC values were high, indicating consistency of the P-Scale, regardless of atypical cases.

The concept of reliability refers to the consistency of assessment scores. It means that an individual is expected to achieve the same result regardless of when the test was applied. The item-by-item reliability for the test-retest showed $\kappa = 83.33\%$, ranging from moderate to almost perfect consistency, indicating acceptable percentages. In this test-retest analysis, only items 2, 15, and 16 showed a coefficient value classified as regular:

(2) Do you work as much as your peers (same number of hours, type of work, etc.)?;

(15) In family discussions, is your opinion important?;

(16) Do you have the same opportunity as your peers to initiate or maintain a lasting relationship with a partner? This result is similar to Souza et al.¹⁴ who showed item reliability of 0.98 when performing an analysis of the P-Scale in terms of item adjustment.

The inter-rater analysis of each item followed the same trend, indicating a high percentage of items with moderate to almost perfect reliability (55.55%). Regular coefficient values were found for items 2, 5, 6, 9, 13, and 16, namely:

(2) Do you work as much as your peers (same number of hours, type of work, etc.)?;

(5) Do you help other people (e.g., neighbors, friends or relatives)?;

(6) Do you participate in recreational/social activities as often as your peers (e.g., sports, conversations, meetings)?;

(9) Do you feel comfortable when you meet new people?;

(13) Do you have the same opportunity to take care of yourself as well as your peers (appearance, nutrition, health)?;

(16) Do you have the same opportunity as your peers to initiate or maintain a lasting relationship with a partner?.

Despite being classified as reasonable, these values are considered acceptable by the literature about inter-rater reliability.²¹

Therefore, the inter-rater reliability of the P-Scale showed good levels of agreement when the items were analyzed separately. Nonetheless, items 11 (Do you move around inside and outside the house and around the neighborhood/city in the same way as your peers?) and 17 (Do you participate as actively

as your peers in religious festivals and rituals (e.g., weddings, baptisms, wakes, etc.)?) reached lower concordance levels, suggesting that, although they are consistent, the assessment of these items showed different results. Souza et al.¹⁴ also found that items 11 and 17 present moderate difficulties for the participants to respond.

In this context, the reproducibility analysis of a measurement should consider that the reliability coefficients are correlated to the stability of repeated measurements, being vulnerable to the influence of different factors.²⁸ For example, the low level of education may hinder the participant self-evaluation of some P-Scale items or even distinguish similar categories. This difficulty is mainly observed in the complexity of the term “peers” used in the P-Scale items, which required different explanations by examiners A and B when providing standardized examples for the participants. Therefore, new strategies should be developed to promote the understanding of the participant or patient during evaluation, especially for items 11 and 17. These issues strengthen the need for more extensive training on the P-Scale to homogenize the measurements, especially for less experienced evaluators.

CONCLUSION

The analysis showed that the P-Scale presented adequate test-retest and inter-rater reliability, with acceptable agreement values. Although items 11 and 17 did not achieve adequate inter-rater reliability, the assessment study was sufficient to guide the correct use of the P-Scale in clinical practice, a scale that proved to be simple, easy, and quick to apply.

These results are meaningful, especially when we recognize the challenges imposed by the complexity of the participation construct and the development of good evaluation tools for measuring it with suitable quality. This complexity makes participation a problematic variable in assessing during rehabilitation. Nonetheless, the results of this study indicate the potential for application of the P-Scale in clinical practice and in scientific research for evaluating the restrictions on participation endured by individuals after stroke.

REFERÊNCIAS

1. Organização Mundial de Saúde. Classificação Internacional de Funcionalidade, Incapacidade e Saúde. São Paulo: EDUSP; 2003.
2. Silva FCM, Sampaio RF, Ferreira FR, Camargos VO, Neves JA. Influence of Context in Social Participation of People with Disabilities in Brazil. *Rev Panam Salud Publica.* 2013;34(4):250-6.
3. Pommerehn J, Delboni MCC, Fedosse E. Classificação Internacional de Funcionalidade, Incapacidade e Saúde e afasia: um estudo da participação social. *CoDAS.* 2016;28(2):132-140. Doi: <https://doi.org/10.1590/2317-1782/201620150102>
4. Aramaki AL, Sampaio RF, Cavalcanti A, Dutra FCMS. Use of client-centered virtual reality in rehabilitation after stroke: a feasibility study. *Arq Neuropsiquiatr.* 2019;77(9):622-31. Doi: <https://doi.org/10.1590/0004-282x20190103>

5. Skoglund E, Westerlind E, Persson HC, Sunnerhagen KS. Self-perceived impact of stroke: A longitudinal comparison between one and five years post-stroke. *J Rehabil Med.* 2019;51(9):660-4. Doi: <https://doi.org/10.2340/16501977-2595>
6. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. *Arch Phys Med Rehabil.* 2002;83(8):1035-42. Doi: <https://doi.org/10.1053/apmr.2002.33984>
7. Gadidi V, Katz-Leurer M, Carmeli E, Bornstein NM. Long-term outcome poststroke: predictors of activity limitation and participation restriction. *Arch Phys Med Rehabil.* 2011;92(11):1802-8. Doi: <https://doi.org/10.1016/j.apmr.2011.06.014>
8. Gretschel D, Visagie S, Inglis G. Community integration of adults with disabilities post discharge from an in-patient rehabilitation unit in the Western Cape. *S Afr J Physiother.* 2017;73(1):361. Doi: <https://doi.org/10.4102/sajp.v73i1.361>
9. Faria-Fortini I, Basílio M, Polese J, Menezes K, Faria C, Scianni A, et al. Caracterização da participação social de indivíduos na fase crônica pós-acidente vascular encefálico. *Rev Ter Ocup USP.* 2017;28(1):71-8. Doi: <https://doi.org/10.11606/issn.2238-6149.v28i1p71-78>
10. Sampaio RF, Silva FCM, Neves JA, Kirkwood RN, Mancini MC. Avaliação dos diferentes domínios do modelo biopsicossocial: uma contribuição para prática clínica utilizando análise fatorial. *Salud(i)Ciencia.* 2013;20(2):134-40.
11. Van Brakel WH, Anderson AM, Mutatkar RK, Bakirtzief Z, Nicholls PG, Raju MS, et al. The Participation Scale: Measuring a key concept in public health. *Disabil Rehabil.* 2006;28(4):193-203. Doi: <https://doi.org/10.1080/09638280500192785>
12. Barbosa, JC; Ramos Jr., AN; Alencar, MJF; Castro, CGJ. Pós-alta em Hanseníase no Ceará: limitação da atividade funcional, consciência de risco e participação social. *Rev Bras Enferm.* 2008;61(esp):727-33. Doi: <https://doi.org/10.1590/S0034-71672008000700012>
13. Dutra FCMS, Mancini MC, Neves JA, Kirkwood RN, Sampaio RF. Empirical Analysis of the International Classification of Functioning, Disability and Health (ICF) Using Structural Equation Modeling. *Braz J Phys Ther.* 2016;20(5):384-394. Doi: <https://doi.org/10.1590/bjpt-rbf.2014.0168>
14. Souza MAP, Coster WJ, Mancini MC, Dutra FCMS, Kramer J, Sampaio RF. Rasch analysis of the participation scale (P-scale): usefulness of the P-scale to a rehabilitation services network. *BMC Public Health.* 2017;17(1):934-42. Doi: <https://doi.org/10.1186/s12889-017-4945-9>
15. Reis BM, Castro SS, Fernandes LFRM. Limitation of activity and restriction of social participation in relation to age range, gender, and education in people with leprosy. *An Bras Dermatol.* 2017;92(3):335-339. Doi: <https://doi.org/10.1590/abd1806-4841.20175216>
16. Dutra FCMS, Prado MC, Borges GPS, Kososki E, Silva FCM. Consequences of pemphigus in occupational performance and social participation of patients. *Salud(i)Ciencia.* 2018;22(8):727-33. Doi: <http://dx.doi.org/10.21840/siic/154617>
17. Van Der Zee CH, Priesterbach AR, Van Der Dussen L, Kap A, Schepers VPM, Visser-Meily JMA, et al. Reproducibility of Three Self-Report Participation Measures: the ICF measure of participation and activities screener, the participation scale, and the utrecht scale for evaluation of rehabilitation-participation. *J Rehabil Med.* 2010;42(8):752-7. Doi: <https://doi.org/10.2340/16501977-0589>
18. Stevelink SA, Terwee CB, Banstola N, Van Brakel WH. Testing the psychometric properties of the Participation Scale in Eastern Nepal. *Qual Life Res.* 2013;22(1):137-44. Doi: <https://doi.org/10.1007/s11136-012-0116-8>
19. Zeeuw J, Douwstra M, Omansen TF, Sopoh GE, Johnson C, Phillips RO, et al. Psychometric Properties of the Participation Scale among Former Buruli Ulcer Patients in Ghana and Benin. *PLoS Negl Trop Dis.* 2014;8(11):e3254. Doi: <https://doi.org/10.1371/journal.pntd.0003254>
20. Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? *J Neurol.* 2012;259(12):2681-94. Doi: <https://doi.org/10.1007/s00415-012-6570-y>
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
23. Munro BH. *Statistical methods for health care research.* 5th ed. Philadelphia: Lippincott, Williams & Wilkins; 2005.
24. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet.* 2014;383(9913):245-254. Doi: [https://doi.org/10.1016/s0140-6736\(13\)61953-4](https://doi.org/10.1016/s0140-6736(13)61953-4)
25. Bensenor IM, Goulart AC, Szwarcwald CL, Vieira MLFP, Malta DC, Lotufo PA. Prevalence of stroke and associated disability in Brazil: National Health Survey - 2013. *Arq Neuro-Psiquiatr.* 2015;73(9):746-50. Doi: <https://doi.org/10.1590/0004-282X20150115>
26. Henriques M, Henriques J, Jacinto J. Acidente Vascular Cerebral no Adulto Jovem: A Realidade num Centro de Reabilitação. *Rev SPMFR.* 2015;27(1):9-13. Doi: <https://doi.org/10.25759/spmfr.180>
27. Malta DC, Felisbino-Mendes MS, Machado IE, Passos VMA, Abreu DMX, Ishitani LH, et al. Fatores de risco relacionados à carga global de doença do Brasil e Unidades Federadas, 2015. *Rev Bras Epidemiol.* 2017;20(Suppl 1):217-32. Doi: <https://doi.org/10.1590/1980-5497201700050018>

28. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59(10):1033-9. Doi: <https://doi.org/10.1016/j.jclinepi.2005.10.015>