






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

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

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

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 Intraclasse (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 ($k_w > 0,81$), 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 ($k_w > 0,81$) a regular ($0,21 < k_w < 0,40$) 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

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 ($k_w > 0.81$), 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 ($k_w > 0.81$) to regular ($0.21 < k_w < 0.40$) 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

¹Universidade Federal do Triângulo Mineiro - UFTM

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

Correspondência

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

Submetido: 4 Junho 2020
Aceito: 30 Dezembro 2021

Como citar

Dutra FCMS, Alves AC, Aramaki AL, Amaral MF, Cavalcanti A. Confiabilidade interavaliadores e teste-reteste da Escala de Participação (P-Scale) em pacientes após acidente vascular cerebral. Acta Fisiatr. 2022;29(1):42-49.



10.11606/issn.2317-0190.v29i1a168523



©2022 by Acta Fisiátrica

Este trabalho está licenciado com uma licença Creative Commons - Atribuição 4.0 Internacional

INTRODUÇÃO

O termo participação refere-se a um estado de saúde e de funcionalidade em que uma pessoa é capaz de se envolver em papéis e situações reais da vida.¹ Participação inclui manter atividades pessoais, mobilidade em diferentes ambientes, relações sociais, trabalho, educação, lazer, espiritualidade e vida comunitária e cível.¹ Este conceito leva em consideração a experiência vivida pelo indivíduo, seu envolvimento e os fatores contextuais que o permeiam, pessoais e ambientais.²

Dessa maneira, participar significa tomar parte, ser incluído ou se envolver em uma área da vida, e ser aceito ou ter acesso aos recursos necessários para esta inclusão.^{1,2}

De acordo com a Organização Mundial de Saúde (OMS), a participação é influenciada por fatores pessoais (como idade, sexo, escolaridade) e fatores ambientais (físicos, sociais e políticos).¹ A interação entre aspectos de saúde, fatores pessoais e fatores ambientais impacta no nível de envolvimento nas atividades diárias e nos papéis socialmente esperados, influenciando a participação.² Restrições em participação são um problema comum entre pessoas com deficiência e em tratamento de reabilitação, como indivíduos após Acidente Vascular Cerebral (AVC).³⁻⁵

O AVC é a principal causa de incapacidade física adquirida em adultos em todo o mundo e representa um problema real de saúde pública em termos de frequência, gravidade e custo. Após o episódio de AVC, cerca de 65% dos indivíduos reportam restrições na reintegração em atividades na comunidade e 54% apresentam alta limitação para realizar atividades de vida diária, tarefas domésticas e compras.⁶ Quando analisados casos crônicos, as restrições em participação se mantêm em pessoas após o AVC. Após quatro anos, quase metade dos sobreviventes de AVC apresentam alguma limitação em atividades e cerca de 30% são classificados como restritos em participação.⁷ As evidências mostram que restrições persistentes na participação têm um impacto significativo na qualidade de vida, no grau de dependência para a realização de atividades de vida diária, na interação social e no desenvolvimento de incapacidades em pessoas com AVC.^{8,9}

A mensuração do grau de restrição em participação pós-AVC é importante para determinar o impacto da incapacidade nesses indivíduos, auxiliar profissionais da saúde a planejar intervenções mais direcionadas a essas demandas, e mensurar a eficácia de intervenções que buscam reduzir as consequências da deficiência. Assim, restrições em participação são informações importantes que devem ser consideradas durante todo o processo de reabilitação em pessoas pós-AVC. A documentação deste construto por meio de instrumentos que identifiquem as necessidades funcionais dos pacientes torna-se relevante, tanto com interesse clínico, como para a pesquisa científica.

Nos países em desenvolvimento, informações sobre a participação de indivíduos com necessidades de reabilitação são limitadas. Neste contexto, a Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF) da Organização Mundial de Saúde (OMS) disponibiliza um modelo conceitual que auxilia o raciocínio clínico e o desenvolvimento de instrumentos de avaliação a partir de uma perspectiva biopsicossocial da saúde.¹⁰ Devido à ampla abordagem e aplicação da CIF, novos instrumentos de avaliação têm sido

desenvolvidos para avaliar e mensurar os domínios e componentes de funcionalidade. Exemplo recente é a Escala de Participação (P-Scale) que avalia a participação em indivíduos com diversas condições de saúde e em diferentes ambientes culturais.¹¹

No Brasil, estudos aplicaram a P-Scale em uma ampla população de pessoas com deficiência atendidas em diferentes serviços de reabilitação.^{2,4,12-16} Um estudo nacional recente testou a adequação dos itens da P-Scale em serviços de reabilitação e forneceu informações úteis para promover uma melhora no instrumento como: demonstrar os itens com mais dificuldade de resposta por parte dos pacientes e propor uma expansão da escala, adicionando itens que se encaixam na restrição dos indivíduos.¹⁴ Na literatura internacional, diferentes estudos apresentam boas propriedades psicométricas e validade da P-Scale,¹⁷⁻¹⁹ porém nenhum estudo avaliou sua confiabilidade no Brasil e de forma específica para pessoas pós-AVC.

OBJETIVO

Avaliar a confiabilidade interavaliador e teste-reteste da versão brasileira da P-Scale em um grupo de pacientes pós-AVC atendidos em um serviço público de reabilitação.

MÉTODO

Estudo metodológico com foco na análise da confiabilidade interexaminador e teste-reteste da P-Scale, aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Triângulo Mineiro (UFTM) (CAAE: 46357215.2.0000.5154).

Seguindo as recomendações de Hobart et al.²⁰ e considerando uma precisão de 0,90, erro estimado de 0,2 e alfa de 0,05, foram selecionados aleatoriamente 20 pacientes com diagnóstico de AVC atendidos em um serviço público de reabilitação. Os critérios para inclusão foram ter acima de 18 anos; ter sofrido AVC há mais de 6 meses e menos de 5 anos; não apresentar outras comorbidades neurológicas e/ou ortopédicas diagnosticadas; e conseguir compreender e responder aos questionários aplicados com fidedignidade.

Descrição da Escala de Participação (P-Scale)

A Escala de Participação (P-Scale) é um instrumento com 18 itens que possibilita a quantificação das restrições na participação percebidas por indivíduos com alguma condição de saúde ou deficiência em oito das nove principais áreas da vida definidas na CIF (aprendizado e aplicação do conhecimento; comunicação e cuidados pessoais; mobilidade; vida doméstica; interação; relacionamentos interpessoais e em comunidade).¹¹

Durante a aplicação da P-Scale, solicita-se que o entrevistado se compare com um 'par' real ou hipotético, alguém que lhe é semelhante em todos os aspectos, exceto pela doença ou deficiência.¹¹ Pergunta-se ao entrevistado se o seu nível de participação é igual ou menor que o de seus pares.

Em seguida, se o entrevistado apontar uma possível dificuldade, ele indica em que grau isto representa um problema em sua vida diária. A pontuação em cada item segue uma escala do tipo likert de 5 pontos em que 0 (zero) indica que o indivíduo não considera sua participação menor do que seu "par"; 1= a participação é menor mas isto não representa um

problema; 2= problema pequeno; 3= problema médio; e 5= problema grande. Para a obtenção do escore final, soma-se os valores de cada item e o valor total do escore da P-Scale varia entre 0 (zero) e 90. Valores menores indicam que o respondente apresenta menos restrições em sua participação.¹¹ O valor assumido como limite máximo para que o indivíduo seja considerado sem restrições na participação é 12 pontos e escores maiores que 12 pontos indicam algum grau de restrição na participação.¹¹ Pacientes que apresentam valores entre 13 e 22 pontos têm restrição leve; valores entre 23 e 32 pontos apresentam restrição moderada; valores entre os limites de 33 a 52 pontos englobam pacientes com restrição severa; e, por fim, valores acima de 53 pontos indicam restrição extrema na participação.¹¹

Procedimentos

Dois avaliadores participaram deste estudo e foram treinados no uso da P-Scale por uma terapeuta ocupacional com experiência no uso do instrumento. O treinamento consistiu em ler o manual traduzido para o português e discutir a forma de aplicação do instrumento. Nenhum dos avaliadores tinha experiência em usar o instrumento nem acesso prévio ao manual. A coleta foi realizada em um serviço público de reabilitação, no período de fevereiro a julho de 2017.

O procedimento de coleta de dados foi realizado em duas etapas. Na primeira etapa, foram coletadas informações sociodemográficas e de saúde como sexo, idade, fatores de risco, auto percepção da saúde, situação de emprego, situação familiar, uso de próteses e/ou órteses e necessidade de ajuda de outras pessoas para executar tarefas diárias; e os participantes responderam à P-Scale sob a forma de entrevista pelo examinador A. Essa primeira etapa durou cerca de 30 minutos. Transcorrido o tempo de avaliação do examinador A, para fins de análise da confiabilidade interexaminador, o participante aguardou por um período de descanso de 30 minutos e o examinador B avaliou os participantes reaplicando a P-Scale, gastando um tempo médio de 20 minutos.

Entre 7 a 10 dias da primeira avaliação ocorreu a segunda etapa dos procedimentos e os participantes foram reavaliados pelo examinador A respondendo a P-Scale. O tempo estimado desta etapa foi de 20 minutos e permitiu analisar o instrumento segundo a confiabilidade teste re-teste. Os examinadores não tiveram acesso aos resultados das avaliações da primeira para que não houvesse memorização dos valores obtidos.

Análise dos dados

A amostra foi caracterizada por meio de estatísticas descritivas. Na análise da confiabilidade teste-reteste e inter-avaliador foram avaliados a confiabilidade dos itens e do escore total da P-Scale. A confiabilidade teste-reteste e inter-avaliador de cada item foi realizada por meio do cálculo do Coeficiente Kappa com ponderação quadrática (kw). Para interpretação do grau de concordância do Kappa adotaram-se os critérios propostos por Landis & Koch:²¹ a) quase perfeita: 0,81 a 1,00; b) forte: 0,61 a 0,80; c) moderada: 0,41 a 0,60; d) regular: 0,21 a 0,40; e) discreta: 0 a 0,20; e) pobre: < 0.

Para avaliação da concordância entre os escores obtidos item por item no teste-reteste e inter-avaliador, foi realizada

análise do gráfico Bland-Altman. O gráfico Bland-Altman calcula a média da diferença entre as duas medidas (teste-reteste ou inter-avaliador) e o intervalo de confiança de 95% para determinar a magnitude da discordância e a existência de erros e padrões sistemáticos entre as duas medidas.²²

O grau de confiabilidade do escore total foi avaliado por meio do Coeficiente de Correlação Intraclasse (ICC) e sua interpretação seguiu os critérios propostos por Munro:²³ muito baixa: 0 a 0,25; baixa: 0,26 a 0,49; moderada: 0,50 a 0,69; alta: 0,70 a 0,89; muito alta: 0,90 a 1,00. Todas as análises foram realizadas no software SPSS IBM® (Statistical Package for the Social Science), versão 20.0.

RESULTADOS

Foram avaliados 20 pacientes, com predominância do sexo masculino (55%) e média de idade de 48,20 anos (DP= 15,68). A maioria (60%) encontra-se casada; 65% residem com cônjuge e filhos e 70% não são responsáveis pela renda familiar. Em relação à escolaridade, a maioria (65%) tem até oito anos de estudo, sendo que 25% possuem ensino fundamental completo. A maioria está afastada do trabalho (55%), recebendo auxílio do Instituto Nacional de Seguridade Social; e seis (30%) estão aposentados.

Em relação aos hábitos de vida, 65% apresentam algum fator de risco como sobrepeso, sedentarismo, consumo de álcool e/ou hábito de fumar; 50% percebem sua saúde física como boa ou muito boa e 65% autoavaliam sua saúde emocional como boa ou muito boa.

A maioria (75%) necessita de auxílio de algum recurso assistivo, com prevalência de 35% de óculos/lentes de contato; e 60% necessitam de algum nível de ajuda para a realização das atividades diárias, sendo 65% auxiliados pelos filhos, cônjuges ou pais. Todos os entrevistados recebem algum tipo de tratamento/medicamento. A Tabela 1 apresenta a descrição completa das informações sociais e de saúde dos participantes deste estudo.

Confiabilidade teste-reteste

Em relação ao escore total da P-Scale, na primeira avaliação, o examinador A encontrou média de 33,45 pontos (DP= 20,74) e variação de 6 a 82 pontos. De forma categórica, 4 (20%) participantes não apresentaram restrições, 2 (10%) apresentaram restrição leve, 4 (20%) foram avaliados com restrição moderada, 7 (35%) com restrição severa, e apenas 3 (15%) participantes foram avaliados com restrição extrema. Já na segunda avaliação, o escore da P-Scale variou de 2 a 80 pontos, com média igual a 30,85 (DP= 21,32). A classificação dos participantes em relação ao grau de restrição apresentou a seguinte categorização: 6 (30%) sem restrição; 3 (15%) restrição leve; 1 (5%) restrição moderada; 7 (35%) restrição severa; e 3 (15%) restrição extrema. Nesta comparação teste-reteste, a análise da concordância do escore total da P-Scale mensurado pelo ICC revelou uma correlação muito alta (ICC= 0,96; IC= 0,89-0,98; p= 0,000) de acordo com os critérios estabelecidos.

Na análise da confiabilidade teste-reteste individualizada para cada um dos 18 itens da P-Scale, a maioria dos itens (3, 4, 5, 10, 11, 13, 14, 17 e 18) apresentaram confiabilidade quase perfeita (kw>0,81); os itens 6, 9, 12 e 17 apresentaram

confiabilidade forte ($0,61 < kw < 0,80$); os itens 1 e 8 tiveram confiabilidade moderada ($0,41 < kw < 0,60$) e confiabilidade regular foi identificada nos itens 2, 15 e 16 ($0,21 < kw < 0,40$). Esta avaliação item por item por meio do coeficiente Kappa com ponderação quadrática (Quadro 1).

A análise gráfica da concordância entre os dois momentos da avaliação (teste-reteste) indicou média das diferenças igual a 1,94 (DP= 8,59), com IC 95% de -2,32 a 6,21. Este valor da diferença das médias se aproxima de zero, indicando que os escores da P-Scale encontrados nos dois momentos da avaliação foram semelhantes, não apresentando erros sistemáticos de mensuração ($p = 0,350$).

Quadro 1. Classificação dos coeficientes Kappa da análise de confiabilidade teste-reteste por itens da P-Scale (n= 20)

Parâmetros de referência para coeficiente de confiabilidade Kappa	Itens
>0,81 (quase perfeita)	3, 4, 5, 10, 11, 13, 14, 17, 18
0,61-0,80 (forte)	6, 7, 9, 12
0,41-0,60 (moderada)	1, 8
0,21-0,40 (regular)	2, 15, 16
0,00-0,20 (discreta)	-
<0 (pobre)	-

O gráfico de concordância de Bland-Altman para o momento teste-reteste (Figura 1). Como pode ser observado, nenhum participante apresentou diferença das médias entre as duas aplicações acima do limite da concordância de 95%, representado pelas linhas tracejadas.

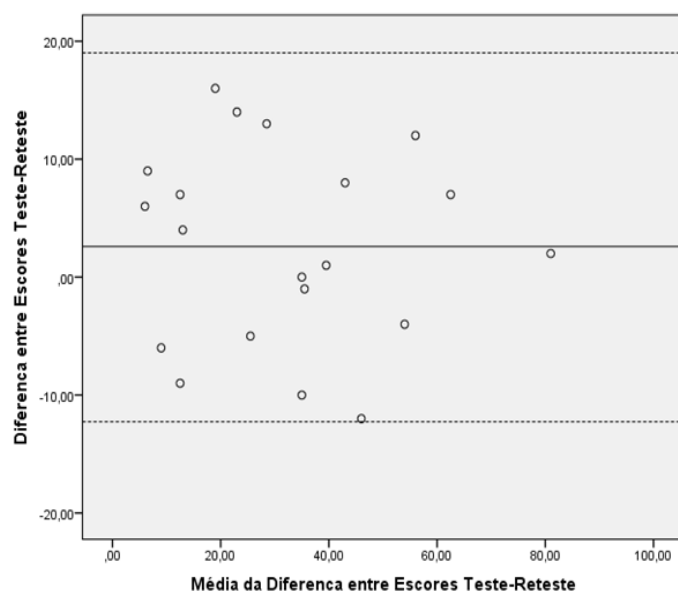


Figura 1. Gráfico de Bland-Altman que examina a concordância teste-reteste da pontuação total da P-Scale (n= 20)

Confiabilidade inter-avaliador

A avaliação item por item da confiabilidade inter-avaliador realizada por meio do coeficiente Kappa com ponderação quadrática (Quadro 2). Das 18 questões da P-Scale, os itens 3, 7, 10 e 15 apresentam confiabilidade quase perfeita ($kw > 0,81$); itens 4, 12 e 18 tiveram confiabilidade forte ($0,61 < kw < 0,80$);

Tabela 1. Informações sociodemográficas e de saúde dos pacientes pós-AVC (n= 20)

Variáveis	Frequência	(%)
Sexo		
Feminino	9	45,00%
Masculino	11	55,00%
Estado Civil		
Solteiro	7	35,00%
Casado	12	60,00%
União Estável	1	5,00%
Com quem Reside		
Sozinho	1	5,00%
Cônjuge	2	10,00%
Filhos	1	5,00%
Pai/Mãe	2	10,00%
Outros Parentes	1	5,00%
Cônjuge e Filhos	13	65,00%
Situação Profissional		
Formal	1	5,00%
Desempregado	2	10,00%
Afastado	11	55,00%
Aposentado	6	30,00%
Nível de Escolaridade		
Não Alfabetizado	2	10,00%
Fundamental I Completo	5	25,00%
Fundamental I Incompleto	2	10,00%
Fundamental II Completo	3	15,00%
Fundamental II Incompleto	1	5,00%
Médio Completo	1	5,00%
Médio Incompleto	5	25,00%
Superior Incompleto	1	5,00%
Principal Responsável pela Renda		
Sim	6	30,00%
Não	14	70,00%
Fatores de Risco		
Fumante	2	10,00%
Sedentarismo	6	30,00%
Fumante e Consumo de Alcool	3	15,00%
Fumante e Sedentarismo	1	5,00%
Consumo de Alcool e Sobrepeso	1	5,00%
Sem Fator de Risco	7	35,00%
Como Avalia Saúde Física		
Muito Boa	1	5,00%
Boa	9	45,00%
Moderada	3	15,00%
Ruim	5	25,00%
Muito Ruim	2	10,00%
Como Avalia Saúde Emocional		
Boa	6	30,00%
Moderada	7	35,00%
Ruim	5	25,00%
Muito Ruim	2	10,00%
Utiliza Algum Recurso Assistivo		
Sim	15	75,00%
Não	5	25,00%
Qual Recurso Utiliza		
Nenhum	5	25,00%
Óculos/Lentes	7	35,00%
Cadeira de Rodas/Banho	3	15,00%
Andador/Bengala/Muleta	1	5,00%
Órteses	1	5,00%
Óculos/Lentes e Aparelhos Auditivos	1	5,00%
Óculos/Lentes e Cadeira de Rodas/Banho	2	10,00%
Recebe Auxílio para Realizar AVD		
Sim	12	60,00%
Não	8	40,00%
Quem Auxilia nas AVD		
Ninguém	7	35,00%
Cônjuge	5	25,00%
Filhos	3	15,00%
Pai/Mãe	5	25,00%
Qual Tratamento Utiliza		
Profissionais de Reabilitação	14	70,00%
Medicamentos em Geral	1	5,00%
Quimioterapia/Radioterapia	1	5,00%
Acompanhamento Médico devido às Queixas da Reabilitação	3	15,00%
Acompanhamento Médico por outras Queixas	1	5,00%

os itens 1, 8 e 14 foram avaliados com confiabilidade moderada ($0,41 < kw < 0,60$); a maioria dos itens (2, 5, 6, 9, 13 e 16) apresentou confiabilidade regular ($0,21 < kw < 0,40$); o item 17 teve confiabilidade discreta ($0,00 < kw < 0,20$) e o item 11 apresentou confiabilidade pobre ($kw < 0,00$).

A análise gráfica da concordância dos itens entre os dois avaliadores indicou média das diferenças igual a $-5,88$ ($DP = 9,58$), com IC 95% de $-10,65$ a $-1,12$. Este valor da diferença das médias se distancia de zero, indicando que os escores encontrados pelos dois avaliadores podem apresentar diferenças sistemáticas de mensuração ($p = 0,018$).

Quadro 2. Classificação dos coeficientes Kappa da análise de confiabilidade inter-avaliador por itens da P-Scale ($n = 20$)

Parâmetros de referência para coeficiente de confiabilidade Kappa	Itens
$>0,81$ (quase perfeita)	3, 7, 10, 15
$0,61-0,80$ (forte)	4, 12, 18
$0,41-0,60$ (moderada)	1, 8, 14
$0,21-0,40$ (regular)	2, 5, 6, 9, 13, 16
$0,00-0,20$ (discreta)	17
<0 (pobre)	11

O gráfico de concordância de Bland-Altman para a avaliação inter-avaliador (Figura 2) indica que dois participantes apresentaram maiores diferenças de escore entre os dois examinadores. Diante desta divergência na concordância, o cálculo do ICC para o escore total da P-Scale foi realizado com a inclusão dos casos atípicos ($n = 20$) e com a exclusão destes participantes ($n = 18$). Assim, os resultados do escore da P-Scale mensurado pelo examinador A (média = $33,45$ pontos; $DP = 20,74$, amplitude de 6 a 82 pontos), foram comparados com os resultados do escore da P-Scale avaliado pelo examinador B que variou de 5 a 85 pontos, com média de $39,05$ pontos ($DP = 22,31$). Nesta comparação da confiabilidade inter-avaliador, o ICC revelou uma correlação muito alta em ambas as análises, sendo $ICC = 0,95$ para 20 indivíduos ($IC = 0,88-0,98$; $p = 0,000$) e $ICC = 0,97$ para 18 casos ($IC = 0,4-0,99$; $p = 0,000$).

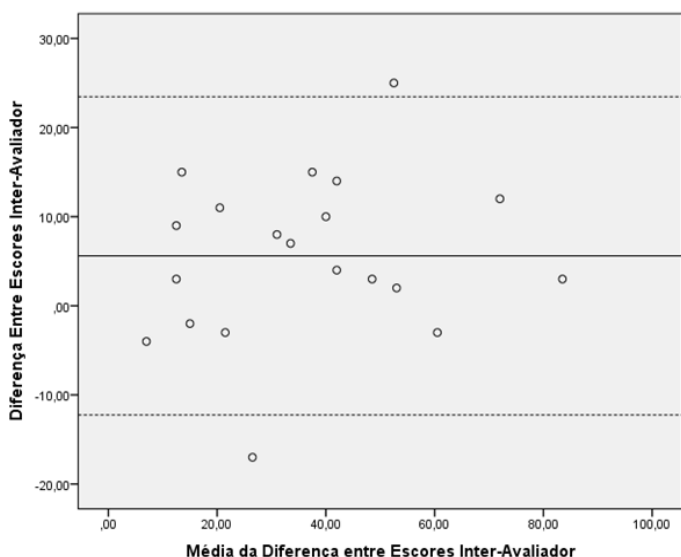


Figura 2. Gráfico de Bland-Altman que examina a concordância interexaminador da pontuação total da P-Scale ($n = 20$)

DISCUSSÃO

Este estudo avaliou a confiabilidade da versão brasileira da P-Scale em um grupo de pacientes pós-AVC atendidos em um serviço público de reabilitação. Os resultados mostraram que a P-Scale apresentou confiabilidade teste-reteste e interexaminador adequadas e com valores de concordância dentro dos limites aceitáveis.

A maioria dos indivíduos avaliados era do sexo masculino, afastado e/ou aposentado e com baixa renda familiar. Este perfil corrobora as características encontradas em outros estudos que avaliaram informações sociodemográficas em pacientes pós AVC.²⁴⁻²⁶ Embora a ocorrência de AVC esteja associada ao envelhecimento, dados de 119 países mostram que o número de casos na faixa etária de 20 a 64 anos aumentou de 25%, em 1990, para 31%, em 2010.²⁴ Esta tendência corrobora com as características dos participantes desta pesquisa que, em sua maioria, são adultos e sofreram o episódio do AVC antes dos 60 anos.

A maioria dos entrevistados relatou hábitos de vida considerados como fatores de risco para o desenvolvimento de doenças crônico-degenerativas, principalmente sedentarismo, um importante fator fortemente correlacionados com a ocorrência de AVC.^{24,25,27} Estes dados seguem as características da população brasileira que apresenta apenas 25% da pessoas ativas fisicamente; média de 13% de fumantes entre homens e 9,3% de mulheres tabagistas; e 14,9% da população com 18 anos ou mais de idade consome bebida alcoólica acima do recomendado pela OMS.²⁷

Apesar destes hábitos de vida seguirem a tendência dos dados nacionais, estes resultados são preocupantes e sugerem uma dificuldade dos pacientes para realizarem mudanças em hábitos considerados fatores de risco associados a doenças crônico-degenerativas. Estas informações são importantes em termos de planejamento de ações de promoção da saúde e de hábitos saudáveis pelos serviços de saúde, principalmente intervenções voltadas à prevenção de doenças junto à população mais vulnerável.

A partir da aplicação da P-Scale, os resultados encontrados pelos dois examinadores e nos dois momentos da avaliação mostraram uma pontuação média que variou de 30 a 39 pontos, aproximadamente. Em relação ao grau de restrição, quando avaliadas conjuntamente, os valores entre os examinadores e entre os dois momentos da avaliação variaram de restrição de moderada a extrema. Outros estudos realizados com população de pessoas com deficiência encontraram valores semelhantes.^{2,13,16}

O ICC é utilizado para avaliar a consistência entre escores provenientes de aplicações repetidas, considerando a variação da amostra.²⁸ Nesta perspectiva, a P-Scale apresentou boa consistência para a amostra estudada, tanto na análise teste-reteste, quando na análise interexaminador.

A confiabilidade interexaminadores apresentou valores adequados para o escore total com correlação de $\alpha = 0,95$, mesmo considerando os casos atípicos. Este valor foi semelhante aos reportados por Stevelink et al.¹⁸ que encontraram índices de confiabilidade interexaminador para a P-Scale de 0,90; e maiores que os resultados encontrados por Van Brakel et al.¹¹ e Van Der Zee et al.¹⁷ que descreveram valores de 0,80 e 0,82, respectivamente.

Estes resultados reportados para a confiabilidade interexaminador indicam que as estratégias adotadas na aplicação da P-Scale foram eficazes para manutenção da consistência das medidas, independente do avaliador. Da mesma forma, o resultado relacionado à confiabilidade teste-reteste mostrou um $\alpha = 0,96$, que representa um valor mais alto de consistência da mensuração quando comparado ao estudo original que desenvolveu a P-Scale.¹¹ Este resultado sugere que o examinador A foi especialmente consistente e os itens da P-Scale são suficientemente claros, contribuindo para a manutenção das medidas sem influência do momento/tempo de aplicação.

O teste de plotagem Bland and Altman mostrou excelente concordância teste-reteste, indicando que os escores obtidos na primeira avaliação estavam concordando com as da segunda em 95% dos casos. Observou-se também excelente concordância interexaminadores, confirmando que os escores do examinador A estavam em concordância com as do examinador B em 95% das avaliações. Este gráfico mostra que não foi observado erro sistemático, ou seja, não houve tendência de subestimar ou superestimar as restrições em participação durante a aplicação da P-Scale.

Entretanto, mesmo diante destes valores, ocorreram dois casos de outliers no diagrama de dispersão, considerando-se a presença de valores nos limites extremos dos intervalos de concordância. Dito de outra forma, dois pacientes, na análise das restrições em participação, apresentaram comportamento atípico, com diferenças entre os escores do examinador A e do examinador B. Desta forma, recomenda-se cautela na interpretação dos dados quando existir uma variabilidade individual elevada, uma vez que esta pode influenciar nos valores médios dos escores obtidos durante as avaliações.²⁸

Para ajustar esta situação, a análise de confiabilidade interexaminador do escore total foi realizada com e sem os indivíduos com comportamento atípico. Em ambas as análises, os valores do ICC foi alto, indicando consistência da P-Scale ao avaliar as restrições na participação, independente dos casos atípicos.

O conceito de confiabilidade refere-se à consistência dos escores de avaliação. Isso significa que é esperado que um indivíduo alcance o mesmo resultado independentemente da ocasião em que este respondeu ao teste. A confiabilidade item por item para o teste-reteste apresentou 83,33% das questões com Kappa ponderado variando de moderado a quase perfeito, o que indica percentuais aceitáveis. Nesta análise teste-reteste apenas os itens 2, 15 e 16 mostraram valor de coeficiente classificado como razoável: (2) Você trabalha tanto quanto os seus pares (mesmo número de horas, tipo de trabalho, etc.); (15) Nas discussões familiares, a sua opinião é importante?; e (16) Você tem a mesma oportunidade que seus pares para iniciar ou manter uma relação duradoura com um companheiro(a)?.

Este resultado se assemelha ao estudo de Souza et al.¹⁴ que apresentou uma confiabilidade dos itens de 0,98 ao realizar uma análise da escala P em termos de ajuste dos itens. A análise inter-examinador de cada item seguiu a mesma tendência, indicando alto percentual de itens com confiabilidade de moderada a quase perfeita (55,55%). Valores de coeficientes razoáveis foram encontrados para os itens 2, 5, 6, 9, 13 e 16, a saber: (2) Você trabalha tanto quanto os seus

pares (mesmo número de horas, tipo de trabalho, etc.); (5) Você ajuda outras pessoas (p. ex., vizinhos, amigos ou parentes)?; (6) Você participa de atividades recreativas/sociais com a mesma frequência que os seus pares (p.ex., esportes, conversas, reuniões)?; (9) Você se sente à vontade quando encontra pessoas novas?; (13) Você tem a mesma oportunidade de se cuidar tão bem quanto seus pares (aparência, nutrição, saúde)?; e (16) Você tem a mesma oportunidade que seus pares para iniciar ou manter uma relação duradoura com um companheiro(a)? Apesar de serem classificados como razoáveis, estes valores são considerados aceitáveis pela literatura que analisa confiabilidade interexaminador.²¹

Assim, a confiabilidade interexaminadores da P-Scale mostrou bons índices de concordância quando analisados item por item. Apesar disso, os itens 11 (Você se locomove dentro e fora de casa e pela vizinhança/cidade do mesmo jeito que os seus pares?) e 17 (Você participa tão ativamente quanto seus pares das festas e rituais religiosos (p.ex., casamentos, batizados, velórios, etc.)) atingiram índices de concordância mais baixos. Isto sugere que, embora sejam consistentes, a avaliação destes itens pelos examinadores apresentou resultados diferentes. Souza et al.¹⁴ ao analisar a dificuldade de resposta aos itens da P-Scale encontraram que os itens 11 e 17 apresentam dificuldade moderada de resposta pelos participantes.

Neste sentido, quando analisamos a reprodutibilidade de uma mensuração, deve-se considerar que os coeficientes de confiabilidade estão relacionados com a estabilidade de medidas repetidas e são vulneráveis à influência de diferentes fatores.²⁸ Como exemplo, a baixa escolaridade da amostra pode ocasionar dificuldades por parte dos participantes em autoavaliarem alguns itens da P-Scale ou em distinguir entre categorias semelhantes. Esta dificuldade é observada principalmente na complexidade do termo “pares” utilizado nos itens da P-Scale, havendo então necessidade de diferenciação na explicação realizada pelos examinadores A e B e fornecimento de exemplos mais padronizados para os participantes. Portanto, novas estratégias devem ser desenvolvidas para facilitar o entendimento do participante/paciente ao ser avaliado, em especial nos itens 11 e 17. Esse resultado reforça a necessidade de treino mais exaustivo no uso da P-Scale para homogeneizar as mensurações, especialmente para avaliadores menos experientes.

CONCLUSÃO

As análises evidenciaram que a P-Scale apresentou confiabilidade teste-reteste e interexaminador adequadas e com valores de concordância aceitáveis. Apesar dos itens 11 e 17 não terem alcançado confiabilidade interexaminador adequada, de forma geral, o estudo do manual foi suficiente no sentido de orientar a correta utilização da P-Scale na prática clínica. Esta mostrou ser simples e de fácil e rápida aplicação.

Estes resultados são importantes, especialmente quando reconhecemos os desafios relacionados à complexidade do construto participação e ao desenvolvimento de bons instrumentos de avaliação capazes de mensurá-lo com qualidade. Essa complexidade torna a participação uma variável difícil de ser avaliada durante a reabilitação. No

entanto, os resultados deste estudo indicam o potencial para aplicação da P-Scale na prática clínica e em pesquisas científicas, com o objetivo de avaliar as restrições na participação apresentadas indivíduos após AVC.

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>