J Musculoskelet Neuronal Interact 2020; 20(4):480-487

Journal of Musculoskeletal and Neuronal Interactions JMNI

Original Article

Inter-rater reliability of physiotherapists using the Action Research Arm Test in chronic stroke

Nicky Spence[1], Nancy C.L. Rodrigues[1], Polykarpos Angelos Nomikos[2], Khalid Mohammed Yaseen[2,3], Mansour Abdullah Alshehri[4]

[1]Sport, Exercise and Physiotherapy Department, School of Health Sciences, University of Salford, United Kingdom;
[2]Academic Rheumatology, School of Medicine, University of Nottingham: Nottingham, Nottinghamshire, United Kingdom;
[3]Physiotherapy Department, School of Rehabilitation, King Abdulaziz University, Jeddah, Saudi Arabia;
[4]Physiotherapy Department, Faculty of Applied Medical Sciences, Umm Al-Qura University, Mecca, Saudi Arabia

Abstract

Objectives: The purpose of this study is to establish whether physiotherapists' ratings are consistent, when using the Action Research Arm Test (ARAT) to score a chronic stroke patient. **Methods**: This was part of a large project establishing the reliability in chronic stroke. This study used a correlational design comparing the association between physiotherapist scores of the same patient, to establish the ARAT's inter-rater reliability. The COSMIN checklist was followed to enhance the methodology of the study. **Results**: Twenty physiotherapists (8 female and 12 male) aged between 25 and 53 years were selected. There were no participant dropouts or withdrawals. The sample size was normally distributed. The physiotherapists appeared representative of the UK physiotherapy population, with the exception of gender. The distribution of scores showed a normal distribution with standard deviation of score of 1.9. The Kendall's W test showed 0.711 of agreement between the raters. The scores achieved statistical significance showing consistency between physiotherapists' scores with chronic stroke. Limitations of the study were the use of a small single center convenience sample that may reduce the generalizability of the findings. **Conclusions**: The ARAT is consistent when scored by physiotherapists in a chronic stroke population. The inter-rater reliability range was (0.70 to 0.90) which is categorized as good.

Keywords: Physiotherapist, Reliability, Stroke

Introduction

Stroke can cause impairments including hemiparesis and spasticity of contralateral upper limbs (UL) which results in long-term disability[1,2]. The costs for the NHS are £3 billion per annum[3]. Stroke is defined by the World Health Organization as "a clinical syndrome consisting of rapidly developing clinical signs of focal or global disturbance of cerebral function lasting more than 24 hours or leading to death with no apparent cause other than a vascular origin." Stroke rehabilitation targets the restoration of function, mobility and movement[5-9] and 60% of patients have persistent UL functional deficits in chronic stage[10,11]. Accurate assessment in stroke's chronic stage is therefore crucial for practice and the selection of appropriate outcome measures is important[12,13]. Equally, rehabilitation in chronic stage of Stroke addresses UL function and an outcome measure that captures these improvements is essential[11,14]. However, suitable outcome measures require adequate psychometric properties within chronic stroke[15].

Many outcome measures have been identified to capture change in UL function. Some of them include the Fugl-Meyer assessment (FMA) scale, the Action Research Arm Test (ARAT), and the Box and Block test (BBT). From the available upper limb outcome measures including the three aforementioned tools all have undergone a mixed degree of psychometric testing[16-18], yet no single item has been tested for all aspects[19]. The BBT is time efficient

The authors have no conflict of interest.

Corresponding author: Mansour Abdullah Alshehri, Abdiya District, Umm Al-Qura University, PO Box 715, Postal Code 21421, Mecca city, Saudi Arabia
E-mail: mamshehri@uqu.edu.sa

Edited by: G. Lyritis
Accepted 4 May 2020

and easy to assess patients however this lacks the detail captured in the ARAT17,20,21. De Weerdt[20] recommend using the ARAT due to its clinical relevance to patients' functional measurement and ease of use. Van der lee et al.[21] reported that the ARAT had greater responsiveness than the FMA scale within chronic stroke. Koh et al.[22] reported the ARAT's validation with stroke. The ARAT is an observer-rated neurological measure used in clinical practice, classified as an Activity measure (level of Domain) according to the International Classification of Functioning, Disability, and Health[23]. The Royal College of Physicians guidelines[24] highlight the need to capture activity and participation for stroke. According to a systematic review, ARAT does not reflect social inclusion and quality of life. Therefore, ARAT users should consider using a measure of participation in conjunction with the ARAT[25].

The ARAT evaluates upper limb function and dexterity following cortical damage[26] adapted from Carrol's[27] Upper Extremity Function Test (UEFT). The Guttman scale was developed within the ARAT to reduce time when performing the tasks[20,28]. Yozbatiran et al.[28] studied a cohort of twelve chronic stroke patients and developed a standardized testing protocol, with the nineteen items tool in order to reduce the discrepancy in scores, as this received criticism in previous studies[29]. The use of the standardised protocol reduces variances between raters, and therefore was selected to be used for this study. Platz et al.[18] and Yozbatiran et al.[28] reported that the ARAT is closely correlated with the FMA establishing the ARAT's construct and concurrent validity. However, a small sample with different diseases was used in Platz's study[18], and only 12 chronic stroke patients participated in Yozbatiran's[28]. The ARAT is a robust test used among stroke, and a range of studies[21,29-31] demonstrated good psychometric properties such as reliability and validity. Two studies[30,32] used occupational therapists to rate ARAT, and one study[31] used physiotherapists, but this was for test retest reliability.

The inter-rater reliability was recorded in three studies as excellent[18,28,33], using Kappa and ICC with small numbers of raters[18]. The studies demonstrated the reliability testing in heterogeneous neurological populations including chronic stroke. Calculating reliability scores between two physiotherapists, further research is required to test the inter-rater reliability with chronic stroke using more than two raters.

This study has the potential to enhance physiotherapists' management of upper limb recovery and indicate the effect of change for both therapist and patient, which may offer encouragement if it shows progression. The transferability of the ARAT findings between staff, if the ARAT is shown to be consistent in the chronic stroke population will enhance the ARAT's utility. The purpose of this study is to establish whether physiotherapists' ratings are consistent, when using the ARAT to score a chronic stroke patient.

Materials and methods

The experimental hypothesis (*H1*) and the null hypothesis (*Ho*) are stated below:

- ARAT scores with chronic stroke are consistent between physiotherapists;
- ARAT scores with chronic stoke are not consistent between physiotherapists and any similarities are due to chance or random error.

Design

The study has a correlational design comparing the association between physiotherapist scores of the same patient, to establish the ARAT's inter rater reliability. The study addressed ethical standards and the Helsinki Declaration revised in 1983 in the University ethics committee approval addressed aspects of risk and benefit covering autonomy, veracity, beneficence and non-maleficence.

The study followed the COSMIN (Consensus-based Standards for the selection of health status Measurements) checklist Box B – reliability testing guidance to enhance the study quality[34]. There was one data collected period for each of the participants, with no intervals. The environment and instructions were the same for all participants. The ARAT weighting scheme was well described addressing COSMIN guidelines.

Equipment

The ARAT questionnaire (Appendix III) and a video recording of a patient undertaking the ARAT was used to ensure consistency of observation. During the ARAT testing participants require a frame for the items either metal or wood, cricket ball, cups, nuts and bars to fit on rods. The video recording was not paused, stopped, fast forwarded, slowed down or repeated in order to accurately reproduce a real life clinical practice environment[35].

Study variables

Dependent variable: ARAT is an observational test which captures upper limb motor function in neurological patients[28] in nineteen tasks addressing four domains: grasp, grip, pinch and gross movement. The scores for each task range from zero (unable to perform the task within 60 seconds), one (partial completion), two (completes task abnormally) to three (performs task within five seconds). The maximum ARAT score is fifty-seven points.

Independent variable: Physiotherapist's ability to interpret the chronic stroke patient video by scoring their upper limb function.

Extraneous variables: Previous papers identified the following confounding factors affecting the scoring of the questionnaire testing environment, noise, temperature, fatigue, mood, distraction[36], size screen and lighting[37,38]. Therefore, all these factors were measured and monitored to ensure study consistency.

Procedure

Convenience and snowball sampling methods were used to recruit the subjects from the University of Salford. Subjects were recruited using posters displayed in common student areas at the University of Salford. Written informed consent was obtained from all participants after initial contact and the information sheet was provided. Twenty subjects were required for the reliability testing based on past reliability studies, as recommended by Steiner and Norman[39]. Lane et al.[40] established minimum sample sizes needed for study reliability in the neurological population and identified that for reliability testing a minimum sample size was twenty participants. All participants were provided with the information sheet and were given 48 hours to deliberate prior to consent. Potential subjects who consented were screened against the study selection criteria to ensure that all were experienced qualified physiotherapists. The researcher provided standardised ARAT instructions to each participant[28]. Demographic and extraneous data were collected using a data extraction sheet. Participants observed the video recording of the chronic stroke patient, and scored it using the standardised ARAT protocol.

Data analysis

The inter-rater reliability of ARAT was analysed comparing the scores of the same patient from twenty different physiotherapists using Kendall's coefficient of Concordance *W*[41-43], as this test is acceptable for between raters in more than three sets of ordinal data[44]. The universally recognised measure of agreement in ordinal data is weighted Kappa. However, weighted Kappa is not a measure of association[45]. Kendall's coefficient of Concordance was selected over weighted kappa as it is a similar measure but has the advantage that it is a measure of association for multiple observers and is useful when twenty observers are utlised in this study[45]. Standard deviation and standard scores were calculated to illustrate the normal distribution of the results. Standard deviation explores how much values differ from the mean scores. Normal distribution was used to illustrate the natural variation of variables around the mean[45]. The results were calculated using the Statistical Package for Social Sciences (SPSS) software package (version 19), using a p value of $p<0.05$, which is typical for therapeutic trials[46-48].

Results

Recruitment and data collection were completed in May 2017. In total, twenty physiotherapists participated in the study; twelve were males and eight were females, with no dropouts recorded. Table 1 represents a heterogenic sample size, which is better to achieve the reliability[39]. A mixed sample is normally seen in the physiotherapy community. Mean age group was 32 years, mode 29 years and median 31 years. The sample represents an international group of physiotherapists with 40% representing EU Physiotherapists and 60% International physiotherapists. Experience ranged from one to thirty-two years. The findings were statistically significant with the p-value being below 0.05. The p-value was found 0.00, and the Kendall's W test statistics showed a good level of agreement (0.71) between the twenty physiotherapists' ratings of ARAT items. The experimental hypothesis was accepted demonstrating agreement between the raters and the ARAT's inter-rater reliability (0.60-0.80)[42]. Table 2 highlights the minimum/maximum score per domain of ARAT so as the mean ranks. Figure 1 compares ARAT scores between raters and the red line shows the mean ARAT scores (17.4 points). The standard deviation (SD) is ±1.9 between the limits of agreement. The range of

Table 1. Participant demographics.

PT	Age	Gender	Country of origin
1	35	M	Saudi Arabia
2	25	M	Greece
3	29	F	Cape Verde
4	46	F	England
5	53	F	England
6	26	M	India
7	33	M	Nigeria
8	34	M	Saudi Arabia
9	31	M	Saudi Arabia
10	29	F	Brazil
11	34	M	Brazil
12	28	F	Brazil
13	36	M	Saudi Arabia
14	27	F	Greece
15	30	F	England
16	27	F	Greece
17	25	M	Greece
18	28	M	Saudi Arabia
19	29	M	Saudi Arabia
20	35	M	Saudi Arabia

Table 2. Mean ranks for ARAT items.

Domains	Questionnaire items	Minimum score	Maximum score	Mean rank
Grasp	Q1 -Q6	5 points	11 points	10.59
Grip	Q7 - Q10	3 points	6 points	11.18
Pinch	Q11 - Q16	1 point	6 points	15.28
Gross movement	Q17 - Q19	4 points	7 points	14.43
Total	Q1- Q19	14 points	22 points	17.4

Table 3. Mean ranks per ARAT item.

ARATs	Mean	ARAT	Mean rank	ARAT	Mean rank
Q1	5.80	Q8	11.53	Q15	6.65
Q2	11.55	Q9	12.53	Q16	9.28
Q3	11.55	Q10	16.00	Q17	12.08
Q4	11.55	Q11	4.25	Q18	12.68
Q5	11.55	Q12	10.15	Q19	18.53
Q6	11.55	Q13	3.45		
Q7	4.68	Q14	4.68		

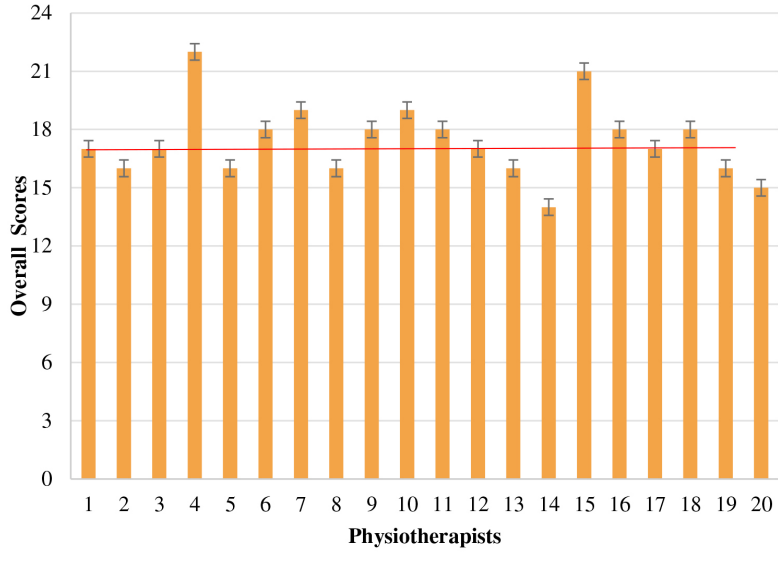

Figure 1. Overall ARAT scores (mean scores between raters).

scores was 22 points (participant 4) to the lowest score of 14 (participant 14). From the SPSS output for the Kendall's W statistics, the mean ranks for the nineteen ARAT items are depicted (Table 3).

Participant's level of Experience

Figure 2 highlights in a pie chart the participants' level of experience. The blue area represents nine (45%) of the sample that had been between one and five years of experience. Within the orange area, seven (35%) of the sample had between five and ten years of experience and within the grey area two (10%) of the sample had between eleven and twenty years of experience. Only two (10%) participants (yellow area) had over 20 years of experience.

Distribution of the results

The distribution of total ARAT scores between raters (Figure 3) peaks between scores of sixteen (n=5), seventeen

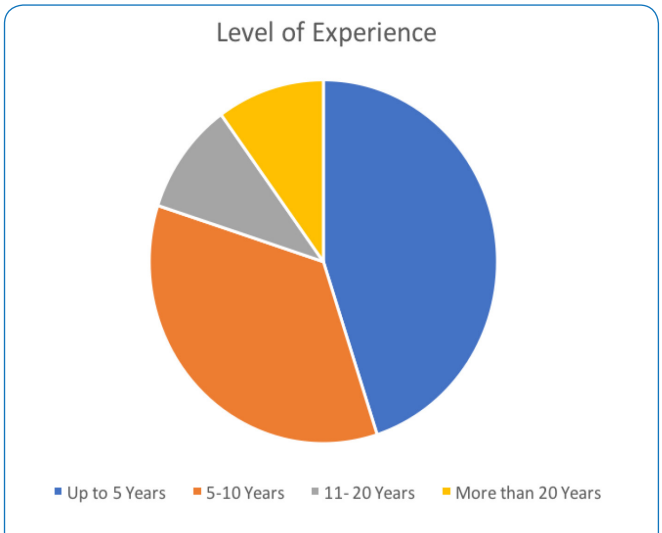

Figure 2. Pie chart participant's level of Experience.

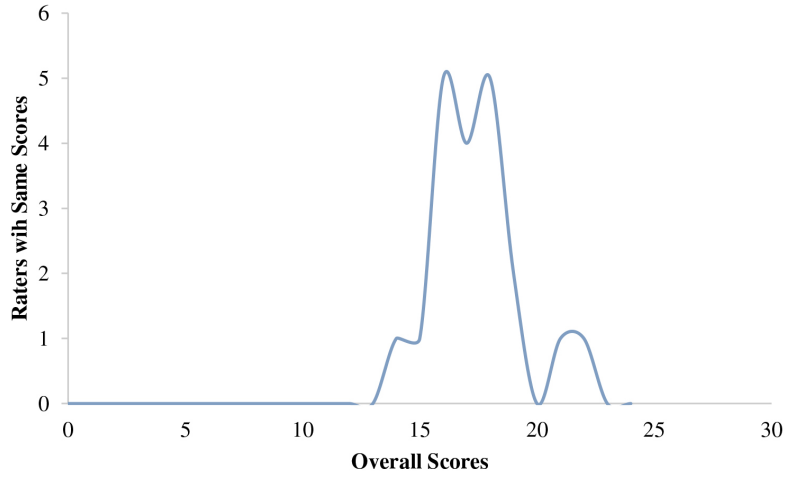

Figure 3. Distribution of rater with the same overall scores.

(n=4) and eighteen (n=5) representing 70% of scores. This illustrates a normal distribution.

Discussion

Modern psychometric testing established that the ordinal outcome measure of upper limb function the ARAT has good inter rater reliability. The statistically significant results at $p<0.05$ indicate that the ARAT was consistent between physiotherapists for chronic Stroke. This represents a good inter-rater reliability (0.70 to 0.90)[42]. The results of the study are supported by similar results achieved by Van der Lee et al[21], Platz et al.[18] and Yozbatiran et al.[28], who achieved excellent scores for inter-rater reliability between two raters. Yozbatirans'[28] and Van der Lees'[33] studies used physiotherapists to rate the ARAT and assess inter-rater reliability in chronic stroke patients. Platzs'[18] study featured a mixed population including stroke, multiple sclerosis and traumatic brain injuries. To reduce type II error, this study utilised a larger sample size using more than two raters when compared to past studies[18,28,33,44].

Stroke guidelines stresses the importance of using psychometrically sound measures when treating stroke. The findings of the study conform to the RCP 5th edition[24], Royal Dutch Society for Physical Therapy (KNGF)[49], Evidence-Based Review for Stroke Rehabilitation (EBRSR)[50], and Scottish Intercollegiate Guidelines Network (SIGN)[19]. The ARAT is explicitly mentioned by both[49,50] stroke guidelines as a measurement tool for upper limb function to be used by physiotherapists. When using the ICF classification the ARAT is an activity measure.

The Figure 3 illustrated the spread of the results around the mean score of 17 and the +/- standard deviation of 1.9 captured the majority of scores showing a limited amount of dispersion about the mean and consistency of rater scores. The distribution of scores was normal. The differing levels of neurological experience (Figure 2) of physiotherapists did not appear to have an adverse effect on the results of this study, as results fell within the 'normal range. The findings of this study are consistent with those of Brunnekreef et al.[38] who also found that experience levels didn't influence reliability. The maximum overall difference in the scores of all the participants was eight points of the total scores, with 22 to 14 being the range, out of a possible 57. Identical scores were achieved in the scoring of items sixteen, seventeen, eighteen and nineteen, representing 70% of the sample (n=14 participants). Two domains showed the greater inconsistencies (Table 2) these were gross movement and pinch.

The ARAT is a useful consistent between tester outcome measure of upper limb function in chronic stroke and it is useful when used between staff, for transferring or discharging patients into a new environment to illustrate present limb function addressing key clinical areas required to inform treatment. The ARAT has its maximum ability when used in moderate degree of stroke severity, due to known ceiling effects. Therefore if chronic stroke participants show mild symptoms it might be worth considering using an alternative due to ARATs inability to detect the smaller milder symptoms.

The use of a standardised protocol enhances reliability by reducing the noise (error), enabling clinicians to capture the patient change (signal). Yozbatiran et al.[28] and future studies[18,29,30,32] have recommended protocol improvements addressing the subjectivity aspects of the ARAT, a consistent patient's posture and scoring informed by performance and time needed per task. If clinicians can use the ARAT in a standardised way, then it will enhance the outcome and their consistency.

Disagreement between raters was observed when sub

scores of zero and one were utilized within the scale individual items. Participants seemed confused when deciding between scores of one and zero. This was not tested though; could be explored by interviewing participants and a greater clarity on what constitutes a zero and one in the standardized testing protocols could be provided. The ARAT test required the patient to complete the task within 60 seconds. Ceiling effects were observed in moderate to chronic stroke patients[21,51]. Chen et al.[51] recommended that ARAT should use a three-point scale combining the zero and one scores with zero to limit confusion.

The majority of UK physiotherapists are female[52], however the study subjects were 60% male and not representative of the UK gender demographics. However, previous ARAT studies did not identify gender as an extraneous factor to influence the results[51]. The participants originated from 7 different countries; with 60% being none EU; that may affect the generalisability to the UK population. Despite the global variation of the physiotherapists[52], this heterogenic sample size has the advantage of achieving better reliability[39], in addition to representing real world clinical practice and experience of physiotherapists. This sample was normally distributed and representative of real world practice for physiotherapy[52].

Known extraneous variables such as mood, tiredness and behaviours, may influence on the results[44], these were not controlled within the study. A further extraneous variable was identified by a participant; the non-use of glasses. It is possible this will have interfered with the data collection, as this participant had the higher score (rater 4). Future studies could monitor glasses use as an extraneous variable and analyse whether glasses wearers completed the test with or without glasses, in order to see if this has an influence on findings.

A difference in screen use when playing the video recording was noted. Fourteen participants observed the video via a projected image, with six participants watching on a laptop screen. The smaller screen could have affected the scoring of the ARAT tasks. Future research could be carried out to analyse whether screen size interferes with the results and is an extraneous variable. However, in this study both groups of participants' scores were in the normal range, suggesting that it did not impact on the findings in this case. Relevant data pertaining to all known and study suggested extraneous variables was captured for post hoc analysis which is beyond the remit of this study highlighting a future area of research.

Conclusion

The ARAT when used with a standardised protocol has demonstrated a statistically significant (0.711) correlation between physiotherapists to a reasonable degree of certainty ($p<0.05$) for treating moderate to severe chronic and should be used in clinical practice today. The study's standard scores showed a normal distribution above and below the mean with the majority of scores within one SD of the mean. The research identified new extraneous variables the wearing/non-wearing of glasses and when video methods are utlised the screen sizes. These factors along with the other known extraneous variables should be monitored when the ARAT is used.

References

1. Sommerfeld DK, Eek EU-B, Svensson A-K, Holmqvist LW, Arbin MHv. Spasticity After Stroke. Stroke 2004; 35(1):134-9.
2. Bleyenheuft Y, Gordon AM. Precision grip in congenital and acquired hemiparesis: similarities in impairments and implications for neurorehabilitation. Front Hum Neurosci 2014;8:459.
3. Stroke Association. State of the nation-stroke statistics. 2018. Available from: https://www.stroke.org.uk/resources/state-nation-stroke-statistics.
4. Force WT. Stroke-1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. Stroke 1989;20(10):1407-31.
5. Dimyan MA, Cohen LG. Neuroplasticity in the context of motor rehabilitation after stroke. Nat Rev Neurol 2011;7(2):76-85.
6. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008;51(1):225-39.
7. Waters KR, Luker KA. Staff perspectives on the role of the nurse in rehabilitation wards for elderly people. J Clin Nurs 1996;5(2):105-14.
8. Nichols-Larsen DS, Clark P, Zeringue A, Greenspan A, Blanton S. Factors influencing stroke survivors' quality of life during subacute recovery. Stroke 2005;36(7):1480-4.
9. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. Stroke 2003;34(9):2181-6.
10. Krakauer JW. Motor learning: its relevance to stroke recovery and neurorehabilitation. Curr Opin Neurol 2006;19(1):84-90.
11. Van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Devillé WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. Stroke 1999;30(11):2369-75.
12. Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. Stroke 2000;31(6):1429-38.
13. Roberts L, Counsell C. Assessment of clinical outcomes in acute stroke trials. Stroke 1998;29(5):986-91.
14. Kwakkel G, Kollen BJ, Wagenaar RC. Therapy impact on functional recovery in stroke rehabilitation: a critical review of the literature. Physiotherapy 1999; 85(7):377-91.
15. Wagenaar RC. Functional recovery after stroke. VU

University Press, 1990.
16. Lin J-H, Hsu M-J, Sheu C-F, Wu T-S, Lin R-T, Chen C-H, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. Phys Ther 2009;89(8):840-50.
17. Lang CE, Bland MD, Bailey RR, Schaefer SY, Birkenmeier RL. Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. J Hand Ther 2013;26(2):104-15.
18. Platz T, Pinkowski C, van Wijck F, Kim I-H, Di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil 2005; 19(4):404-11.
19. Smith L. Management of Patients With Stroke: Rehabilitation, Prevention and Management of Complications, and Discharge Planning: a National Clinical Guideline. SIGN, 2010.
20. De Weerdt W, Harrison M. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. Physiother Can 1985;37(2):65-70.
21. van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. Improving the Action Research Arm test: a unidimensional hierarchical scale. Clin Rehabil 2002;16(6):646-53.
22. Koh C-L, Hsueh I, Wang W-C, Sheu C-F, Yu T-Y, Wang C-H, et al. Validation of the action research arm test using item response theory in patients after stroke. J Rehabil Med 2006;38(6):375-80.
23. World Health Organization. International classification of functioning disability and health. Geneva, 2001.
24. Bowen A, James M, Young G. Royal College of Physicians 2016 National clinical guideline for stroke. RCP, 2016.
25. Sivan M, O'Connor RJ, Makower S, Levesley M, Bhakta B. Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. J Rehabil Med 2011;43(3):181-9.
26. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res 1981;4(4):483-92.
27. Carroll D. A quantitative test of upper extremity function. J Chronic Dis 1965;18(5):479-91.
28. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. Neurorehabil Neural Repair 2008;22(1):78-90.
29. Van Der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. J Rehabil Med 2001;33(3):110-3.
30. Hsueh I-P, Lee M-M, Hsieh C-L. The Action Research Arm Test: is it necessary for patients being tested to sit at a standardized table? Clin Rehabil 2002; 16(4):382-8.
31. Nomikos PA, Spence N, Alshehri MA. Test-retest reliability of physiotherapists using the action research arm test in chronic stroke. J Phys Ther Sci 2018; 30(10):1271-7.
32. Hsieh C-L, Hsueh I-P, Chiang F-M, Lin P-H. Inter-rater reliability and validity of the action research arm test in stroke patients. Age Ageing 1998;27(2):107-13.
33. Van der Lee J, De G. V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra-and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. Arch Phys Med Rehabil 2001;82(1):14-9.
34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual. Amsterdam, University Medical Center, 2012.
35. Shrum W, Duque R, Brown T. Digital Video as Research Practice: Methodology for the Millennium. J Res Pract 2005;1(1):M4.
36. Asan O, Montague E. Using video-based observation research methods in primary care health encounters to evaluate complex interactions. Inform Prim Care 2014;21(4):161.
37. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped observational gait-analysis assessments. Phys Ther 1991;71(6):465-72.
38. Brunnekreef JJ, Van Uden CJ, van Moorsel S, Kooloos JG. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. BMC Musculoskelet Disord 2005;6(1):17.
39. Streiner DL, Norman GR, Health measurement scales: a practical guide to their development and use. Oxford University Press, USA; 1995.
40. 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.
41. Kendall MG, Smith BB. The problem of m rankings. Ann Math Stat 1939;10(3):275-87.
42. Legendre P. Species associations: the Kendall coefficient of concordance revisited. J Agr Biol Envir St 2005;10(2):226.
43. Salkind N. Encyclopaedia of research design. Sage, 2010.
44. Hicks C. Research for Physiotherapists: Project Design and Analysis. Churchill Livingstone, 1995.
45. Norman GR, Streiner DL. PDQ statistics. PMPH-USA, 2003.
46. Feys HM, De Weerdt WJ, Selz BE, Cox Steck GA, Spichiger R, Vereeck LE, et al. Effect of a therapeutic intervention for the hemiplegic upper limb in the acute phase after stroke: a single-blind, randomized, controlled multicenter trial. Stroke 1998;29(4):785-92.
47. Hammond JA. Assessment of clinical components of physiotherapy undergraduate education: are there any issues with gender? Physiotherapy 2009; 95(4):266-72.
48. Gross R, Kinnison N. Psychology for nurses and health professionals. Routledge, 2017.

49. Dutch Society of Neuro Rehabilitation. KNGF Clinical Practice Guideline for Physical Therapy in patients with stroke. Nederlands, 2014.
50. Iruthayarajah J, Mirkowski M, Reg MMO, Iliescu A, Caughlin S, Harris J, et al. Upper Extremity Motor Rehabiliation Interventions. Available from: http://www.ebrsr.com/evidence-review/10-upper-extremity-interventions.
51. Chen H-f, Lin K-c, Wu C-y, Chen C-l. Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. Arch Phys Med Rehabil 2012;93(6):1039-45.
52. Sykes C. In perspective: Window on the world - physiotherapy differs from country to country. The Chartered Society of Physiotherapy 2014. Available from: https://www.csp.org.uk/frontline/article/perspective-window-world-physiotherapy-differs-country-country.