Published in final edited form as:

OTJR (Thorofare N J). 2019 January ; 39(1): 64–73. doi:10.1177/1539449218757740.

Interpreting Action Research Arm Test assessment scores to plan treatment

Emily S. Grattan, PhD, OTR/L,

Department of Health Sciences & Research, Medical University of South Carolina

Division of Occupational Therapy, Medical University of South Carolina

Craig A. Velozo, PhD, OTR/L, FAOTA,

Division of Occupational Therapy, Medical University of South Carolina

Elizabeth R. Skidmore, PhD, OTR/L, FAOTA, FACRM,

Department of Occupational Therapy, University of Pittsburgh

Stephen J. Page, PhD, OTR/L, MS, FAHA, FAOTA, FACRM, and

Division of Occupational Therapy, The Ohio State University

Michelle L. Woodbury, PhD, OTR/L

Department of Health Sciences & Research, Medical University of South Carolina, Division of

Occupational Therapy, Medical University of South Carolina, Ralph H. Johnson VA Medical

Center, Charleston, SC

Abstract

Background—Rasch keyforms can help interpret clinical assessment scores. The Action Research Arm Test (ARAT) is a commonly used assessment, yet no keyform currently exists.

Objective—To provide a keyform for the ARAT and demonstrate how a clinician can use the keyform to design optimally challenging rehabilitation sessions.

Methods—Secondary analysis of ARAT data (n=122) using confirmatory factor and Rasch analyses to examine the measurement properties and generate a keyform.

Results—The Item standardized factor loadings were >0.40 (range 0.82-0.96) and R-square values were >0.60 (range 0.65-0.96). All items exhibited adequate infit statistics with point measure correlations >0.60 (range 0.72-0.97). Person reliability was 0.98 and person separation was 7.07. Item difficulty measures ranged from -2.78 logits to 2.64 logits.

Correspondence: Emily S. Grattan, PhD, OTR/L, Department of Health Science and Research, 77 President St. MSC 700, Medical University of South Carolina, Charleston, SC 29425, Phone: 843-792-3435, Fax: 843-792-1358; grattan@muscc.edu.

Conflict of Interest: The authors declare that there is no conflict of interest (Grattan, Velozo, Skidmore, Page & Woodbury).

Research Ethics and Patient Consent: All research procedures were approved by the Medical University of South Carolina Institutional Review Board (PRO00013941); University of Pittsburgh Institutional Review Board (PRO07070003; PRO07110071; PRO12040650; PRO11110467); The Ohio State University Institutional Review Board PRO2011H0216. Interpreting Action Research Arm Test assessment scores to plan treatment

Conclusion—The ARAT has strong measurement properties, and a keyform was provided. We showed how the keyform can be utilized by clinicians to interpret scores, set goals, and plan treatment.

Keywords

stroke; rehabilitation; upper extremity; Rasch analysis; assessment

Approximately 85% of stroke survivors in the acute stage of stroke exhibit upper extremity (UE) hemiparesis and 55-75% stroke survivors in the chronic stage of stroke experience UE functional limitations (Mozaffarian et al., 2016; Wolf et al., 2006). Thus, UE motor recovery is a priority rehabilitation goal (Barker & Brauer, 2005; Bohannon, Andrews, & Smith, 1988). Rehabilitation professional practice guidelines (American Occupational Therapy Association, 2014; American Physical Therapy Association, 2014) direct therapists to address impairments and limitations identified by standardized assessments. However, standardized assessments are infrequently used (Abrams et al., 2005; Kay, Myers, & Huijbregts, 2001; Menon-Nair, Korner-Bitensky, Wood-Dauphinee, & Robertson, 2006; Swinkels, van Peppen, Wittink, Custers, & Beurskens, 2011; Torenbeek, Caulfield, Garrett, & Van Harten, 2001) in part because of therapists' difficulty interpreting scores (Jette, Halbert, Iverson, Miceli, & Shah, 2009; Swinkels et al., 2011).VELOZO and Woodbury (2011) argued that an assessment score, derived by summing ordinal item ratings, fails to inform clinical decisions because it does not detail specific behaviors that a client performs entirely or in part. The growing emphasis on clinical use of standardized assessments necessitates development of more clinically-interpretable scoring methods to inform care-planning.

The Action Research Arm Test (ARAT) (Lyle, 1981) is a reliable, valid (Lang, Wagner, Dromerick, & Edwards, 2006; Platz et al., 2005) standardized assessment of post-stroke hemiparetic UE functional limitation. The instrument contains 19 items grouped into 4 subtests; grasp, grip, pinch, and gross motor. Items within the subtests are arranged in an item-difficulty order originally proposed by Lyle et al. (1981). The assessment can be administered in two ways; testing only the items relevant to the client's ability as per original instructions (Lyle, 1981) or testing all items (van der Lee, Roorda, Beckerman, Lankhorst, & Bouter, 2002; Yozbatiran, Der-Yeghiaian, & Cramer, 2008). Item performance is rated on a 4-point scale (0=unable; 1=partial; 2= abnormal; 3=normal) then item ratings are summed and reported out of 57 points with higher score indicating greater UE function.

The ARAT aggregate score has limited clinical interpretability because it does not indicate which items were easy, difficult, or optimally challenging for the client. For example, a score of 50 indicates less limitation than a score of 20, but does not specify if the client could grip cylindrical objects or pinch small objects. A therapist needs detailed behavioral information to tailor therapy to clients' specific abilities. One way to gain this detailed information is to examine both the aggregate score and the client's pattern of item responses.

Rasch analysis offers a framework and method to increase an assessment's clinical interpretability with an output called a "keyform." A keyform is a pencil-paper scoring template (Kielhofner, Dobria, Forsyth, & Basu, 2005; Linacre, 1997;VELOZO, Warren, Hicks, & Berger, 2013) upon which a therapist records item ratings then examines the pattern of

item responses. The keyform is formatted according to Rasch model expectations that a client will successfully accomplish easy items, have less success with difficult items, and have a 50% probability of success on items with a difficulty level similar to his/her ability level. The item response pattern will be consistent with this expectation; the client will have a consistent pattern of good ratings on easy items, a consistent pattern of poor ratings on difficult items, and a fluctuation of ratings on a region of the keyform called the transition zone (Veloze & Woodbury, 2011) where item difficulties match the clients' ability level. The transition zone indicates behaviors (tested by assessment items) that are optimally difficult for an individual (Linacre, 1997). This information is important because a key ingredient of motor recovery therapy is task-specific practice (TSP) at the optimal, just-right, level of challenge (Guadagnoli & Lee, 2004). Therefore, the keyform allows a clinician to identify appropriately challenging behaviors, represented by items in the transition zone, to target in TSP sessions (Bode, Heinemann, Kozlowski, & Pretz, 2014; Kielhofner et al., 2005; Veloze & Woodbury, 2011; Woodbury, Veloze, Richards, & Duncan, 2013). Rasch keyforms have been used to improve the scoring, interpretation and clinical utility of assessments of vision (Veloze et al., 2013), disorders of consciousness (Pape, Mallinson, & Guernon, 2014), functional independence (Bode et al., 2014), pediatric gross motor function (Avery, Russell, Raina, Walter, & Rosenbaum, 2003), and post-stroke UE impairment (Woodbury et al., 2016). To our knowledge, there is only one UE stroke assessment (Veloze & Woodbury, 2011; Woodbury et al., 2016) with a keyform and there are no keyforms for measuring post-stroke UE function. Given the dearth of rehabilitation keyforms, few clinicians are aware that they exist or understand how they are used.

This paper, using the ARAT as an exemplar, demonstrates how keyforms may inform care planning. The purposes are to generate an ARAT keyform and demonstrate how a clinician can use it to design optimally challenging rehabilitation sessions. Several groups have applied Rasch or other item response analysis methods to the ARAT (Chen, Lin, Wu, & Chen, 2012; Koh et al., 2006; van der Lee et al., 2002) and found that the rating scale was adequate and that the ARAT is a precise and reliable measure. Therefore, we believe the ARAT is suitable for keyform development.

METHODS

Participants and Study Design

A secondary analysis of existing data pooled from five stroke rehabilitation intervention studies conducted at three academic health/research centers was performed. Study sites were located in mid-sized cities in the United States which provided a diverse sample.

The full 19-item ARAT (i.e., all items tested) was administered at each site by trained therapists using similar standardized procedures. De-identified pre-intervention data were obtained from each site in accordance with local data sharing regulations. It was feasible to pool data because the studies had similar eligibility criteria. Participants were included if they experienced a stroke 3 months prior and had UE hemiparesis. Participants were excluded if they had severe hemiparesis (e.g., no palpable triceps contraction); severe spasticity (Modified Ashworth Scale >3 in elbow, wrist, or fingers); were unable to follow one-step commands (due to severe aphasia or cognitive impairment); or reported UE pain.

All procedures were approved by local institutional review boards and adhered to the ethical standards of the revised Declaration of Helsinki. All participants or their proxies provided informed consent.

Data Analysis

Confirmatory Factor Analysis—Confirmatory factor analysis (CFA) tested the hypothesis that all items measure a single skill (unidimensionality) (Kline, 2005; Velozo & Woodbury, 2011). Four models were fit; 1-factor, 2-factor (Factor 1: grasp, grip, pinch; Factor 2: gross motor) 3-factor (Factor 1: grasp, grip; Factor 2: pinch; Factor 3: gross motor), and 4-factor (Factor 1: grasp; Factor 2: grip; Factor 3: pinch; Factor 4: gross motor) (M-Plus version 6)(Muthén & Muthén, 2010) using weighted least-squares means with variance adjusted (Brown, 2006). Model fit was verified with the comparative fit index (CFI), Tucker Lewis index (TLI) and Root Mean Square Error of Approximation (RMSEA). The CFI and TLI compare the hypothesized model to a null model; values >0.95 indicate fit. The RMSEA compares actual to expected model parameters; values <0.06 indicate fit. The RMSEA is influenced by sample size (Kenny, Kaniskan, & McCoach, 2014) therefore, each item's standardized factor loadings and R-square values were also examined to indicate the level of association between the item and underlying construct; factor loadings values >0.40 and item R-square values >0.60 indicate adequate association (Hu & Bentler, 1999; Kline, 2005; Marsh, Balla, & McDonald, 1988). The fit indices for each model were compared to determine the one that best fit the data.

Rasch Analysis—The Rasch rating scale model (Winsteps v.3.75) (Bond & Fox, 2007; Wright & Stone, 1979) was applied to examine ARAT measurement properties (Linacre, 2012).

Rating Scale—The rating scale was examined to assure that lower ratings were consistently given to lower functioning individuals and vice versa. Following Linacre's guidelines (Linacre, 2002), adequate rating scale diagnostics were defined as; >10 observations per rating category, average measures advance monotonically with each category, category thresholds increase with each category, and the outfit mean square residual value for each category <2.0.

Item Fit—Item infit statistics, reported as mean square residuals (MnSq) with associated standardized z-values, indicated how well the data fit the Rasch model (Wright & Stone, 1979), specifically with regards to items that closely match subjects' ability (Wright, Linacre, Gustafson, & Martin-Lof, 1994). Clinical observation MnSq values <0.5 or >1.7 with Z score >2.0 =misfit (Wright et al., 1994).

Point Measure Correlations—Point measure correlations indicated the degree to which each item represented the underlying construct: <0.30=weak, 0.31-0.59=moderate, >0.60=strong (Andresen, 2000).

Person Reliability and Separation—Person reliability, analogous to Cronbach's alpha, and person separation indicated how well the items differentiated ability levels. The person

separation index was used to calculate the number of statistically distinct strata (i.e., ability levels) into which the assessment divided subjects using the equation: $(4 \times \text{Separation index} + 1)/3$ (Fisher, 1992; Wright & Masters, 1982). We considered reliability values >0.90 , separation index values >2 , and strata ≥ 3 as adequate for this clinical application (Wright & Masters, 1982).

Item Difficulty Hierarchy—Rasch analysis calculated item-difficulty and person-ability measures to a single metric (logit units which are equal unit intervals) so that items were ordered from easy-to-difficult and the sample was ordered from less-to-more ability. We examined item-person map to ascertain the congruency between the Rasch derived and originally proposed (Lyle, 1981) item difficulty hierarchies.

Keyform and its Use for Goal Setting and Treatment Planning—An ARAT keyform was generated from the “general keyforms” output option (Winsteps v.3.75) (Linacre, 1997). To illustrate the process of using the keyform to interpret the ARAT score, we randomly selected a participant with high ability (ARAT score 39-57), displayed his/her raw data on the keyform, identified a transition zone, described appropriately challenging therapy goals, and provided examples of how the goals could link to treatment.

RESULTS

Demographics

The pooled $n=122$ dataset yielded a diverse sample with a wide range of ability (Table 1). Participants averaged 57 years of age ($SD=13.7$) with ischemic stroke and moderate UE impairment (ARAT total score, $mean=27.3$, $SD=16.6$).

Confirmatory Factor Analysis

The results for the 1-factor, 2-factor, 3-factor, and 4-factor analyses are presented in Table 2. All models had similar fit indices, thus we defined the most parsimonious 1-factor model as best fitting the data (Brown, 2006). The 1-factor CFI and TLI both met criterion ($CFI=0.99$, $TLI=0.99$). The RMSEA did not meet criterion ($RMSEA=0.15$), however all standardized estimates and R-square values exceeded criterion (estimates all >0.4 , range 0.82-0.98; and R-square >0.6 , range 0.65-0.96) thus supporting essential unidimensionality (Kline, 2005).

Rating Scale

For each rating category: there were >10 observations, average measures increased monotonically, the category thresholds increased with each category, and the outfit MnSq values <2.0 (Table 3).

Item Fit

All items exhibited adequate infit statistics (Table 4).

Point Measure Correlations

All point measure correlations were >0.60 (Table 4).

Person Reliability and Separation

Person reliability =0.98, person separation =7.07, and the subjects were divided into 9 statistically distinct levels.

Item Difficulty Hierarchy

Item difficulty measures (Table 4) ranged from -2.78 logits (hand to mouth) to 2.64 logits (ball bearing ring & thumb). Within each subtest, item difficulties were congruent with the original item difficulty order. For example, in the grasp subscale, the 2.5 cm block was the easiest and the 10 cm block was the most difficult item. Some items had similar item difficulty measures. For example, the washer and pouring water items from the grip subtest and the marble index and thumb item from the pinch subtest had similar difficulty levels (0.24, 0.32 and 0.38 logits respectively). The item-person map (Figure 1) illustrates that the items were well-matched to the range of the samples' ability measures (i.e., mean of item difficulties was 0.0 and mean of person abilities was 0.13).

Keyform

Figure 2 presents an ARAT keyform for an individual with high ability (ARAT score=42). On the figure's right side, items are arranged in descending difficulty order from hard (top) to easy (bottom). On the figure's left side, the 4-point rating scale is plotted relative to the measurement metric at the figure's base. The ratings stair-step upwards from left to right as item difficulties increase. The participant's actual item ratings are circled. Consistent with Rasch model expectations, this person performed easy items well (consistent ratings =3 on easy items) but had difficulty with harder items (consistent ratings =2 on the majority of hard items). In the middle, the ratings deviate back and forth between adjacent ratings, e.g., 3 to/from 2. This region, the transition zone, indicates behaviors (tested by specific items) for which the participant has some, but not full, ability to perform. Conceptually this zone represents the point at which the patient is transitioning from one level ability to the next higher level of ability and mathematically has a 50% probability of receiving adjacent item ratings. Behaviors tested by items in the transition zone are expected to recover sooner than behaviors tested by items above the transition zone. Therefore, transition zone items (e.g. cricket ball, pour water) suggest grasp patterns (e.g., spherical grasp, cylindrical grasp) that can be addressed in shorter-term therapy goals. Items above the transition zone (e.g. ball bearing) suggest prehension patterns (e.g., palmar prehension) that can be addressed in longer-term goals. A second keyform is also provided as an example of an individual with moderate ability (Figure 2).

Keyform for Treatment Planning

The goals suggested by the keyform transition zone can be linked to treatment activities. For example, the keyform (Figure 2) indicated that the therapist address the prehension patterns associated with the "cricket ball" (spherical grasp), "pouring water," (cylindrical grasp), and "marble index and thumb" (pincher grasp) for short term goals. Table 5 links this information to possible therapy activities. For example, a therapist may set a functional goal (e.g. self-feeding) that requires use of a pincher grasp and then select and grade repetitive task-specific practice activities that promote practice and mastery of this prehension pattern.

DISCUSSION

This study aimed to generate an ARAT keyform and demonstrate its use for setting goals and planning treatment. First, we explored the ARAT's item-level measurement properties to assure that it accurately measured UE function. While others have conducted factor analyses, item response and Rasch analyses of the ARAT, the unique contribution of this paper is the ARAT keyform. The ARAT keyform enables clinicians to link patient evaluation to the design of an individualized treatment plan.

Dimensionality

We found that the full 19-item ARAT is unidimensional, a result consistent with the literature (Chen et al., 2012; Koh et al., 2006; van der Lee et al., 2002). Lyle (1981) suggested the assessment be scored as 4 sub-tests. Our analysis supports current practice which is to sum all items and report an aggregate score.

Rating scale

The 4-point rating scale was adequate in this sample, a finding inconsistent with Chen et al. (2012) who suggested combining categories 0 and 1 and Koh et al. (2006) who combined categories 1 and 2. Differences in sample characteristics may explain the discrepancies. For example, Chen's (2012) sample had higher and less varied ARAT scores (mean=35.78, SD=16.7) suggesting they may have not had enough lower-ability participants to adequately evaluate lower rating categories. Similarly, Koh's (2006) sample had lower ARAT scores (median=5.0, IQR=0-40) thus may not have had enough higher ability participants to evaluate higher ratings.

Item Fit

We showed that all items fit the Rasch model. In contrast, Chen et al. (2012) found 2 items (hand behind head, hand to top of head), Koh et al. (2006) found 3 items (ball bearing ring/thumb, marble ring/thumb, 10 cm block) and van der Lee et al. (2002) found 4 items (ball bearing ring/thumb, ball bearing middle/thumb, ball bearing index/thumb, marble ring/thumb) that failed to fit the model tested. In our study the ball bearing ring to thumb item approached misfit which, when taken together with the literature, may indicate "noise" in this item. One interpretation is that factors other than the construct being measured influence the response to a misfitting item. Perhaps biomechanical or anatomical factors influence a person's ability to perform this prehension pattern more so than prehension patterns tested by other items.

Person Reliability and Separation

Similar to the body of literature (Chen et al., 2012), we found the ARAT to be reliable and precise. The person separation value (7.07) indicated the ARAT is able to differentiate the sample's ability into 9 strata. In contrast, Chen et al. (2012) reported a lower value (3.83) and differentiation into 4 strata. The difference in results is likely because our sample had a wider range of ARAT scores, thus more ability strata. However, the clinical relevance of being able to detect 9 versus 4 ability strata remains unknown and could be explored in future studies.

Item Difficulty Hierarchy

The Rasch-derived item difficulty order was consistent with Lyle's originally hypothesized order, a finding consistent with van der Lee et al. (2002). However, our results differ from Chen et al. (2012) and Koh et al. (2006) who found that pinch (Chen) and grip (Koh) subtest items were not consistent with Lyle's order. Interestingly, our results indicate that several items had similar difficulty measures. van der Lee et al. (2002) removed 4 items having similar difficulty measures to improve the assessment's clinical efficiency. However, from a clinical perspective similarly difficult items may provide important information. For example, two grip items and one pinch item had similar difficulty measures. But, item removal should be approached with caution because a therapist may require information about both gross motor and pinch skills in order to comprehensively understand the person's ability.

Keyform

Koh et al. (2006) discussed the importance of establishing the clinical usefulness of the ARAT score. We approached the issue of clinical interpretability by linking the ARAT's qualitative content (the behaviors tested by each item) to its quantitative aspects (item ratings) via a new scoring method, the keyform. Our findings confirm the ARAT has strong measurement properties, which made it suitable for a keyform. Previous studies presented keyforms for other rehabilitation assessments and demonstrated their application to clinical practice (Bode et al., 2014; Pretz et al., 2015; Velozo et al., 2013; Velozo & Woodbury, 2011). The ARAT keyform now joins this growing family of keyforms available to therapists.

We illustrated how a clinician can use the ARAT keyform to design treatment. As shown in the examples, a therapist could use a keyform to plan treatment by first administering the ARAT and circling the patient's item ratings on the keyform. Next, the therapist would locate the transition zone by following the consistent pattern of ratings at the bottom of the keyform upwards until it deviates to the next lower adjacent rating (e.g. from a rating of 3 to 2 or from a rating of 2 to 1, or from a rating of 1 to 0). This deviation will mark the lower boundary of the transition zone which is defined as the first 6 consecutive items for which at least 4 of these items received the next lowest rating. These 6 items represent the expected next steps in the subject's transition from a current skill level to a greater skill level and will therefore be the movements to target in the task practice therapy session. For example, the participant (higher ability) represented in Figure 2 received high ratings on easy items, lower ratings on more difficult items, and had fluctuating ability on moderately difficult items evidenced as a region of back-and-forth ratings in the middle of the keyform. This region is the transition zone because it represents the client's transition from one ability level to the next higher ability level. ARAT items within this zone are neither too easy nor too difficult and indicate different grasp/prehension patterns at the "just right" challenge level. The therapist can then use this information to plan treatment because the transition zone displays which items a client is more likely to improve on in the short term thus forming the basis for therapy goals and intervention approaches. The goals and activities presented are examples of ways a clinician could use this information to personalize goals and treatment activities

that are meaningful to each client. The ARAT keyform offers a unique and novel tool that clinicians can use to facilitate treatment planning which may enhance patient outcomes.

Study Limitations

A limitation of this study was the relatively small sample size and corresponding lack of statistical power used to conduct the CFA. In addition, keyforms are less useful for individuals of very high/low ability because there is greater measurement error at these extremes due to floor/ceiling effects (Veloze & Woodbury, 2011). Clinicians may find that using a keyform with these individuals is less effective however, these individuals are not being measured well by the ARAT regardless of whether there is a keyform. The keyform presented highlights potential short term/long term goals. However, the demarcation is inexact and provides a general guideline to illustrate how goals can be established using the keyform.

Future Studies

While we believe that the ARAT keyform is a useful tool for clinicians, future studies should examine the feasibility of using the keyform in clinical practice to determine whether the keyform has an impact on clinicians' use of the ARAT. Future studies should also examine whether using the ARAT keyform enhances UE outcomes.

Implications for Occupational Therapy Practice

Our findings indicate that the ARAT has strong measurement properties and provides a logical basis to generate a keyform. The keyform may facilitate occupational therapists use of the ARAT in clinical practice due to its applicability in goal setting and treatment planning.

Acknowledgments

Supported by: National Institutes of Health (UL1RR024153, UL1TR000005, R21 AT002110-01, R01 AT004454-05); University of Pittsburgh, School of Health and Rehabilitation Science Research Development Fund; University of Pittsburgh's Office of Research Central Research Development Fund; Ralph H. Johnson VA Medical Center and the Office of Research Development, Rehabilitation Research and Development, Department of Veterans Affairs; VA Merit Review Award (NO799-R); Allergan, Inc.

References

- Abrams D, Davidson M, Harrick J, Harcourt P, Zylinski M, Clancy J. 2005; Monitoring the change: Current trends in outcome measure usage in physiotherapy. *Manual Therapy*. 11(1):46–53.
- American Occupational Therapy Association. 2014; Occupational Therapy Practice Framework: Domain and Process (3rd edition). *American Journal of Occupational Therapy*. 68:S1–S48.
- Andresen EM. 2000; Criteria for assessing the tools of disability outcomes research. *Archives of physical medicine and rehabilitation*. 81(Supplement 2):S15–S20. [PubMed: 11128900]
- Avery LM, Russell DJ, Raina PS, Walter SD, Rosenbaum PL. 2003; Rasch analysis of the Gross Motor Function Measure: validating the assumptions of the Rasch model to create an interval-level measure. *Archives of Physical Medicine & Rehabilitation*. 84(5):697–705. [PubMed: 12736885]
- Barker RN, Brauer SG. 2005; Upper limb recovery after stroke: The stroke survivors' perspective. *Disability and Rehabilitation*. 27(20):1213–1223. DOI: 10.1080/09638280500075717 [PubMed: 16298923]

- Bode RK, Heinemann AW, Kozlowski AJ, Pretz CR. 2014; Self-scoring templates for motor and cognitive subscales of the FIM instrument for persons with spinal cord injury. *Arch Phys Med Rehabil*. 95(4):676–679.e675. DOI: 10.1016/j.apmr.2013.11.009 [PubMed: 24309071]
- Bohannon RW, Andrews AW, Smith MB. 1988; Rehabilitation goals of patients with hemiplegia. *International Journal of Rehabilitation Research*. 11(2):181–184.
- Bond, TG, Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. 2nd. Routledge; 2007.
- Brown, TA. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.
- Chen HF, Lin KC, Wu CY, Chen CL. 2012; Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. *Arch Phys Med Rehabil*. 93(6):1039–1045. DOI: 10.1016/j.apmr.2011.11.033 [PubMed: 22420887]
- Fisher WP. 1992; Reliability statistics. *Rasch measurement transactions*. 6(3):238.
- Guadagnoli MA, Lee TD. 2004; Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav*. 36(2):212–224. DOI: 10.3200/jmbr.36.2.212-224 [PubMed: 15130871]
- Guide to Physical Therapy Practice 3.0. Alexandria, VA: American Physical Therapy Association; 2014.
- Hu, Lt; Bentler, PM. 1999; Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*. 6(1):1–55.
- Jette DU, Halbert J, Iverson C, Miceli E, Shah P. 2009; Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther*. 89(2):125–135. [PubMed: 19074618]
- Kay TM, Myers AM, Huijbregts MPJ. 2001; How far have we come since 1992? A comparative survey of physiotherapists' use of outcome measures. *Physiotherapy Canada*. 53:268–275.
- Kenny DA, Kaniskan B, McCoach DB. 2014 The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*.
- Kielhofner G, Dobria L, Forsyth K, Basu S. 2005; The construction of keyforms for obtaining instantaneous measures from the occupational performance history interview rating scales. *OTJR: Occupation, Participation and Health*. 25(1):23–32.
- Kline, RB. Principles and Practice of Structural Equation Modeling 2005. New York, NY: Guilford; 2005.
- Koh CL, Hsueh IP, Wang WC, Sheu CF, Yu TY, Wang CH, Hsieh CL. 2006; Validation of the action research arm test using item response theory in patients after stroke. *J Rehabil Med*. 38(6):375–380. DOI: 10.1080/16501970600803252 [PubMed: 17067971]
- Lang C, Wagner J, Dromerick A, Edwards D. 2006; Measurement of upper-extremity function early after stroke: Properties of the Action Research Arm Test. *Archives of physical medicine and rehabilitation*. 87:1605–1610. [PubMed: 17141640]
- Linacre J. 1997; Instantaneous measurement and diagnosis. *Physical Medicine and Rehabilitation*. 11:315–324.
- Linacre, J. A user's guide to Winsteps and Ministeps Rasch-model computer programs. Chicago, IL: WINSTEPS; 2012.
- Linacre JM. 2002; Optimizing rating scale category effectiveness. *J Appl Meas*. 3(1):85–106. [PubMed: 11997586]
- Lyle RC. 1981; A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *International Journal of Rehabilitation Research*. 4(4):483–492. [PubMed: 7333761]
- Marsh HW, Balla JR, McDonald RP. 1988; Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*. 103(3):391–410.
- Menon-Nair A, Korner-Bitensky N, Wood-Dauphinee S, Robertson E. 2006; Assessment of unilateral spatial neglect post stroke in Canadian acute care hospitals: are we neglecting neglect? *Clinical rehabilitation*. 20(7):623–634. [PubMed: 16894806]

- Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Turner MB. 2016; Executive Summary: Heart Disease and Stroke Statistics—2016 Update: A Report From the American Heart Association. *Circulation*. 133(4):447–454. DOI: 10.1161/cir.0000000000000366 [PubMed: 26811276]
- Muthén, B, Muthén, L. Mplus (Version 6). Los Angeles, CA: Muthen & Muthen; 2010.
- Pape TL, Mallinson T, Guernon A. 2014; Psychometric properties of the disorders of consciousness scale. *Archives of physical medicine and rehabilitation*. 95(9):1672–1684. DOI: 10.1016/j.apmr.2014.04.015 [PubMed: 24814459]
- Platz T, Pinkowski C, van Wijck F, Kim I, di Bella P, G J. 2005; 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*. 19:404–411. [PubMed: 15929509]
- Pretz CR, Kean J, Heinemann A, Kozlowski AJ, Bode R, Gebhardt E. 2015; A Multidimensional Rasch Analysis of the Functional Independence Measure based on the NIDILRR Traumatic Brain Injury Model Systems National Database. *J Neurotrauma*. doi: 10.1089/neu.2015.4138
- Swinkels RA, van Peppen RP, Wittink H, Custers JW, Beurskens AJ. 2011; Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. *BMC Musculoskelet Disord*. 12:106.doi: 10.1186/1471-2474-12-106 [PubMed: 21600045]
- Torenbeek M, Caulfield B, Garrett M, Van Harten W. 2001; Current use of outcome measures for stroke and low back pain rehabilitation in five European countries: first results of the ACROSS project. *International Journal of Rehabilitation Research*. 24(2):95–101. [PubMed: 11421397]
- van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. 2002; Improving the Action Research Arm test: a unidimensional hierarchical scale. *Clin Rehabil*. 16(6):646–653. [PubMed: 12392340]
- Veloza CA, Warren M, Hicks E, Berger KA. 2013; Generating clinical outputs for self-reports of visual functioning. *Optom Vis Sci*. 90(8):765–775. DOI: 10.1097/oxp.0000000000000007 [PubMed: 23839700]
- Veloza CA, Woodbury ML. 2011; Translating measurement findings into rehabilitation practice: An example using the Fugl-Meyer Assessment of the Upper Extremity with clients following stroke. *Journal of Rehabilitation Research and Development*. 48(10):1211–1222. [PubMed: 22234665]
- Wolf SL, Winstein CJ, Miller JP, Taub E, Uswatte G, Morris D, Investigators, E. 2006; Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA*. 296(17):2095–2104. DOI: 10.1001/jama.296.17.2095 [PubMed: 17077374]
- Woodbury ML, Anderson K, Finetto C, Fortune A, Dellenbach B, Grattan E, Hutchison S. 2016; Matching Task Difficulty to Patient Ability During Task Practice Improves Upper Extremity Motor Skill After Stroke: A Proof-of-Concept Study. *Arch Phys Med Rehabil*. 97(11):1863–1871. DOI: 10.1016/j.apmr.2016.03.022 [PubMed: 27117385]
- Woodbury ML, Veloza CA, Richards LG, Duncan PW. 2013; Rasch Analysis Staging Methodology to Classify Upper Extremity Movement Impairment After Stroke. *Archives of physical medicine and rehabilitation*. 94(8):1527–1533. DOI: 10.1016/j.apmr.2013.03.007 [PubMed: 23529144]
- Wright BD, Linacre JM, Gustafson J, Martin-Lof P. 1994; Reasonable mean-square fit values. *Rasch measurement transactions*. 8(3):370.
- Wright, BD, Masters, GN. Rating Scale Analysis. Rasch Measurement. ERIC; 1982.
- Wright BD, Stone MH. 1979 Best Test Design. Rasch Measurement.
- Yozbatiran N, Der-Yeghiaian L, Cramer SC. 2008; A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair*. 22(1):78–90. DOI: 10.1177/1545968307305353 [PubMed: 17704352]

Figure 1.

ARAT 19 Person-Item Map. Person abilities (X) (left side of figure) and item difficulties (right side of figure) are displayed linearly on a logit (equal unit intervals) scale. Persons of high ability and difficult items are located at the top of the figure. Persons of low ability and easy items are located on the bottom of the figure. A wide distribution (–6 to 7 logits) of person abilities and item difficulties are present. Item difficulties are mean item difficulties rather than the spread of the rating scale threshold. The mean person abilities and item difficulties are similar indicating that item difficulties are well matched to person abilities. M, mean; S, 1 SD; T, 2 SDs.

Figure 2.
Keyform for subject with high ability and subject with moderate ability.

Table 1

Demographic and descriptive data

	Sample (n=122)	Study 1 (n=24)	Study 2 (n=20)	Study 3 (n=20)	Study 4 (n=23)	Study 5 (n=35)
Age, years, <i>M</i> (<i>SD</i>)	57.2 (13.7)	54.4(10.4)	66.7 (11.7)	59.6 (12.0)	46.7 (14.0)	59.2 (12.9)
Female, <i>n</i> (%)	63 (51.6)	13 (54.2)	10 (50.0)	10 (50.0)	15 (65.2)	15 (42.9)
Caucasian, <i>n</i> (%)	89 (73.0)	12 (50.0)	14 (70.0)	16 (80.0)	21 (91.3)	26 (74.3)
Stroke Etiology						
Ischemic, <i>n</i> (%)	75 (61.5)	14 (58.3)	17 (85.0)	8 (40.0)	16 (69.5)	20 (57.1)
Stroke Hemisphere						
Right, <i>n</i> (%)	50 (41.0)	7 (29.2)	11 (55.0)	8 (40.0)	14 (69.9)	10 (28.6)
Dominance						
Right, <i>n</i> (%)	111 (91.0)	21 (87.5)	17 (85.0)	19 (95.0)	22 (95.7)	32 (91.4)
Dominant side affected, <i>n</i> (%)	60 (49.2)	14 (66.7)	7 (53.8)	9 (45.0)	10 (43.5)	20 (57.1)
ARAT at Baseline						
<i>M</i> (<i>SD</i>)	27.3 (16.6)	25.9 (19.7)	31.2 (15.7)	29.7 (15.9)	23.0 (14.8)	27.6 (16.3)
Range	1-56	1-56	3-55	1-54	4-54	4-54
Low [*] , <i>n</i> (%)	45 (36.9)	10 (41.7)	5 (25.0)	5 (25.0)	12 (52.2)	13 (37.2)
Moderate [†] , <i>n</i> (%)	44 (36.1)	8 (33.3)	8 (40.0)	8 (40)	7 (30.4)	13 (37.2)
High [‡] , <i>n</i> (%)	33 (27.0)	6 (25.0)	7 (35.0)	7 (35.0)	4 (17.4)	9 (25.7)

Note. ARAT=Action Research Arm Test; ARAT score range, 0-57;

^{*} Low ability=scores 0-19;

[†] Moderate ability=scores 20-38;

[‡] High ability=scores 39-57; Ability level cut-offs defined by authors for descriptive purposes.

Table 2

Confirmatory Factor Analyses

	CFI	TLI	RMSEA	Chi-Square	Model fit	Item Standardized Factor Loadings (range)	R-square values (range)
1-Factor	0.989	0.987	0.145		0.000	0.815-0.979	0.664-0.958
2-Factor	0.993	0.992	0.116		0.000	0.893-0.980	0.797-0.961
3-Factor	0.994	0.994	0.102		0.000	0.904-0.981	0.817-0.978
4-Factor	0.995	0.994	0.099		0.000	0.909-0.990	0.881-0.981

Note. CFI=Comparative fit index; TLI=Tucker Lewis index; RMSEA=Root mean square error of approximation.

Table 3

Rating scale diagnostics based on Linacre's criteria

Category	Observed Count (n=1935)	Percentage of Counts	Average Logit Measure	Outfit	Rasch-Andrich Threshold
0 (cannot perform)	647	28	-3.54	1.26	None
1 (partially performed)	363	16	-1.58	0.64	-1.97
2 (performed abnormally)	860	37	0.87	1.02	-1.32
3 (performs normally)	427	19	4.36	1.09	3.29
Missing	21	1	-2.31	-	-

Table 4

Item fit statistics and item difficulty hierarchy by subtest

Item	Standardized Factor Loadings	Infit MnSq	Infit Zstd	Point measure correlation	Originally proposed hierarchy for subtest	Item Difficulty Measure	SE of the difficulty measure
Grasp Items							
Block 2.5 cm	0.94	0.68	-2.6	0.83	Easiest	-1.32	0.17
Block 5 cm	0.98	0.55	-3.8	0.87		-0.89	0.17
Cricket ball	0.95	0.58	-3.4	0.88		-0.44	0.17
Stone	0.95	0.78	-1.6	0.86		-0.35	0.17
Block 7.5 cm	0.97	0.72	-2.1	0.89		-0.20	0.17
Block 10 cm	0.95	0.77	-1.6	0.88	Most difficult	0.51	0.17
Grip Items							
Tube 2.25 cm	0.93	0.81	-1.4	0.96	Easiest	-1.39	0.17
Tube 1 cm	0.94	0.74	-1.9	0.84		-1.16	0.17
Washer	0.93	0.73	-2.0	0.88		0.24	0.17
Pour water	0.89	0.84	-1.1	0.86	Most difficult	0.32	0.17
Pinch Items							
Marble index & thumb	0.94	1.09	0.6	0.88	Easiest	0.38	0.17
Marble middle & thumb	0.95	0.92	-0.5	0.89		1.05	0.18
Marble ring & thumb	0.93	1.00	0.1	0.84		1.36	0.18
Ball bearing index & thumb	0.93	1.19	1.2	0.94		1.64	0.18
Ball bearing middle & thumb	0.96	1.26	1.6	0.82		2.10	0.18
Ball bearing ring & thumb	0.92	1.67	3.7	0.87	Most difficult	2.64	0.19
Gross Movement Items							
Hand to mouth	0.74	1.15	1.0	0.97		-2.78	0.19
Hand to top of head	0.88	1.35	2.3	0.72		-1.20	0.17
Hand behind head	0.82	1.53	3.2	0.72	Most difficult	-0.5	0.17

Table 5

Linking ARAT scores to treatment

ARAT Item	Prehension or grasp pattern	Example of Goal	Possible Treatment Activities
Cricket ball	Spherical grasp	In two weeks, client will be able to use paretic UE to pick up and toss ball 90% of trials to participate in leisure activities.	Repetitive task-specific practice using paretic UE to pick up objects of various sizes and weights (e.g. ball, orange), use paretic UE to stabilize container, practice using computer mouse; open lever handle doors
Pouring water	Cylindrical grasp	In three weeks, client will be able to prepare simple breakfast (cereal, juice) using paretic UE to pour cereal, milk, and juice without spilling contents.	Repetitive task-specific practice using paretic UE to pick up objects of various sizes, weights, and contents (e.g. can, jar cup, water, sugar) and pour contents into container; water flowers; use hammer
Marble, index and thumb	Pincher grasp	In three weeks, the client will be able to use paretic UE to self-feed small finger foods 80% of trials.	Repetitive task-specific practice using paretic UE to pick up/manipulate objects of various shapes and sizes (e.g. fasten buttons, place coins in pouch, play leisure game with small game pieces)