

Investigating Psychometric Properties of the Arm Activity Measure – Thai Version (ArmA-TH) Subscales Using the Rasch Model

Montana Buntragulpoontawee (✉ montana.b@cmu.ac.th)

Department of Rehabilitation Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, 50200, Thailand. <https://orcid.org/0000-0003-3388-7720>

Jeeranan Khunachiva

Department of Rehabilitation Medicine, Faculty of Medicine, Chiang Mai University

Patreeya Euawongyarti

Department of Rehabilitation Medicine, Faculty of Medicine, Chiang Mai University

Nahathai Wongpakaran

Department of Psychiatry, Faculty of Medicine, Chiang Mai University

Tinakon Wongpakaran

Department of Psychiatry, Chiang Mai University Faculty of Medicine

Atcharee Kaewma

Department of Rehabilitation Medicine, Nakhornping Hospital

Stephen Ashford

Regional Hyper acute Rehabilitation Unit, London, North West University Healthcare NHS Trust, Centre for Nursing, Midwifery and Allied health Research, University College London Hospitals, University College London & Department of Palliative Care, Policy and Rehabilitation, King's College London

Research

Keywords: ArmA-TH, Rasch analysis, psychometric properties, upper limb hemiplegia, outcome measure

Posted Date: November 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-100177/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Research Methodology on March 9th, 2021. See the published version at <https://doi.org/10.1186/s12874-021-01238-5>.

Abstract

Background: This study investigated the ArmA-TH measurement properties based on item response theory, using the Rasch model.

Methods: Patients with upper limb hemiplegia resulting from cerebrovascular and other brain disorders were asked to complete the ArmA-TH questionnaire. Rasch analysis was performed to test how well the ArmA-TH passive and active function sub-scales fit the Rasch model by investigating unidimensionality, response category functioning, reliability of the person and item, and differential item functioning (DIF) for age, sex and education.

Results: Participants had stroke or other acquired brain injury ($n=185$) and the majority were men 126(68.1%), with a mean age of 55(SD 22). Most patients 91(49.2%) graduated elementary/primary school. For the ArmA-TH passive function scale, all items had acceptable fit statistics. The scale's unidimensionality, and local independence were supported. The reliability was acceptable. Disordered threshold was found in five items, none was DIF. For the ArmA-TH active function scale, one item was misfitting and three were locally dependent. The reliability was good. DIF was not found. All items had disordered thresholds, and data fitted the Rasch model better after rescoreing.

Conclusions: Both sub-scales of ArmA-TH fitted the Rasch model, and are valid and reliable. The disordered thresholds should be further investigated.

Background

The Arm Activity Measure questionnaire (ArmA) is a twenty-item patient and/or carer-reported outcome measure of function of the hemiparetic upper limb developed in 2013 by Ashford et al., with primary goal to address "real-life" function, that is, day-to-day performance in the person's normal environment [1]. The unique characteristic of ArmA is its two separate constructs, having passive and active function sub-scales, in order to evaluate the most clinically relevant goals [2]. As there has never been an objective self-report measure to assess hemiparetic upper limb function for patients in Thailand, ArmA was translated into the Thai language as ArmA-TH with preliminary psychometric properties evaluation including content validity index both for item (I-CVI) and score (S-CVI), inter-rater reliability and internal consistency [3]. The construct validity of the items and detailed item evaluation of ArmA-TH based on measurement theory, however, was not initially explored. According to measurement theory, an outcome measure scale should demonstrate that all items contribute to the same construct, has invariance across a sub-population [4]. All these properties can be evaluated by conformity to the Rasch model, which the original English version of the ArmA passive function sub-scale has been evaluated against in a UK sample [5].

In this study, we, therefore, aimed to examine the extent to which our data, from a Thai sample, fit the Rasch measurement model. The Rasch model belongs to the item-response latent trait models, a probabilistic logistic model that predicts that the response to a particular item is influenced by the quality of both person and item. The key concepts of Rasch model are; first, transforming non-linear raw scores

into logit scale measures, location (logit) of both particular person and item are determined on the same interval scale. This interval scale can differentiate how people adhered to the fundamental measurement principle, which provides interval level measurement as apposed to ordinal scaling using the raw score [6]. Second, the “invariance of parameters” [4, 6], which implies that the calibration of the functionality of ArmA-TH is sample distribution free and the calibration of persons is indicator distribution free along the arm function continuum. Put other words, the location of any two persons on the continuum should be independent of the items used to make that comparison [7]. For the purpose of achieving Rasch measurement model, it requires that “data fit the model” [7], in contrast to other item response theory (IRT) models and classic test theory (CTT), in which “the model is modified to fit the data” [6].

Methods

Population

The authors asked patients with hemiplegic upper limb impairment resulting from stroke and other acquired brain injury visiting rehabilitation services in Chiang Mai, completed the ArmA-TH questionnaire.

All patients were between 20-85 years of age with Thai as their mother tongue and graduated from at least elementary school with the ability to understand Thai communication in daily activities. The patients’ demographic characteristics recorded for this study were age, sex, hemiparetic side, diagnosis, education level and ArmA-TH passive and active scores.

Measure

The ArmA-TH is a twenty-item questionnaire for assessing difficulty in functioning of hemiparetic upper limb. There are seven items in the passive function sub-scale and thirteen items in the active function sub-scale. Using a Likert scoring system between 0 (no difficulty) and 4 (unable to do task), the passive function sub-scale scores range from 0 (high function) to 28 and the active function sub-scale scores range from 0 (high function) to 52 [2].

Analysis

Descriptive statistics are used to describe demographic characteristics of the patients, presenting as mean (SD). The ArmA-TH sub-scale scores presented in median (inter-quartile range).

Rasch analysis

To test whether the data fit the Rasch model, the following criteria were investigated [8, 9].

1. Unidimensionality. Two methods were evaluated for determining unidimensionality. First, the first principal component of the residuals (first construct) should be no more than 15% or Eigen value less than 2 [10]. Second, that the item fit statistics indicating the extent to which the response to a particular item are consistent with the way the sample respondents have responded to the other items, should be 0.70 and 1.50 [10]. In addition, the correlation of the two sets of person measures and the correlation disattenuated for measurement error, should be greater than 0.7 to indicate unidimensionality.
2. Local independence. To evaluate local independency, a pair of items should not have inter-item residual correlations that higher than 0.2 [11].
3. Response category functioning. Ordered categories and thresholds are expected for measurement. Therefore, adjacent categories (thresholds) on the latent scale hold the same position and order on the latent trait measured [12]. Items with a disordered threshold between categories can be evaluated by category probability curves and the item fit of each categorical response is examined (less than 2.0, are acceptable) [8].
4. Targeting of persons, items and item hierarchy. Acceptable item-test targeting for compliance with the Rasch model is evaluated through the closeness of the mean of person and the mean of item on the Wright map (no more than 1 logit) [13]. Item hierarchy indicates how the items in difficulty match the intentions of the instrument developer and the expectations of those planning to use the test results [14].
5. Reliability. There are two kinds of reliability evaluated by Rasch analysis, the person and the item reliability. The person reliability is interpreted as the ability of the scale to reliably rank the person relative to location within the scale of the measure. Similar to Cronbach's alpha, the value is often lower because it does not include extreme scores. The item reliability coefficient reflects the extent to which the item hierarchy is replicable with a different set of individuals. A reliability coefficient of > 0.70 is considered acceptable for person, and coefficient of ≥ 0.80 is considered acceptable for item.

Differential item functioning (DIF) for age, sex and education. An ideal item is that it should be invariant across subgroups, meaning that item calibration should be the same in different subgroups of people [8]. Moderate to large DIF was evaluated by significant DIF Contrast' of < 0.64, thereby indicating an acceptable value [6]. In this study, DIF due to age, sex and education was examined. Both the ArmA-TH passive and active function sub-scales were separately evaluated for fit to the Rasch model.

Winsteps, 4.5.3 (Winsteps® Rasch Measurement, 2017) was used for Rasch analysis.

Results

A total of 185 patients participated in the questionnaire evaluation. The majority were men 126 (68.1%) with mean age of 55 (SD 22). Hemiparesis resulted from hemorrhagic stroke 81 (43.8%), ischemic stroke 78 (42.1%), traumatic brain injury 24 (13.0%) and other causes 2 (1.1%). Most patients 91 (49.2%) graduated at elementary/primary school level, followed by secondary school level 40 (21.6%), vocational

or high vocational certificate 28 (15.1%), and the least at level of bachelor degree and above 26 (14.1%). The ArmA-TH passive function sub-scale scores range from 0 to 28, covering the total range from minimum to maximum score. The ArmA-TH active function sub-scale scores range from minimum of 0 to 49, almost reach the maximum score of 52. Details are shown in Table 1.

Table 1
Demographic characteristics and ArmA-TH sub-scale scores of 185 patients completing the ArmA-TH questionnaire.

Demographic characteristics	Number (%) (<i>n</i> = 185)	
Mean age (years)	55	(SD 22.0)
Sex	126	(68.1)
Male	59	(31.9)
Female		
Diagnosis	81	(43.8)
Hemorrhagic stroke	78	(42.1)
Ischemic stroke	24	(13.0)
Traumatic brain injury	2	(1.1)
Other brain injury		
Education	91	(49.2)
Primary school	40	(21.6)
Secondary school	28	(15.1)
Vocational or high vocational certificate	26	(14.1)
Bachelor degree and above		
ArmA	Median (Inter-quartile range)	
Passive Function Sub-scale	6	(2–11)
Active Function Sub-scale	11	(5–18)

Analysis according to the Rasch model

For the ArmA-TH passive function, the fit statistics ranged from 0.73 to 1.31 indicating all items contributed to the Rasch measurement model. Principal component analysis of residuals showed the 1st Eigen value of 1.73 (11.5%) supporting the unidimensionality, whereas the standardized residual correlations were less than 0.3 indicating local independence. The person reliability was acceptable (0.70), while the Cronbach's alpha was 0.83. The item reliability was excellent (0.97). No disordered category was found, however, disordered threshold was found in item 1 (Cleaning Palm), item 2 (Cutting

finger nails), item 3 (Putting on a glove), item 6 (Put on a splint), and item 7 (Positioning arm on a cushion or support in sitting) (Table 2). The ArmA-TH passive function seemed not to be well-targeted in this sample as the mean logit between item and person was more than 1 SD. Item bias or DIF was not found in ArmA-TH passive function.

Table 2
Rasch analysis results of. ArmA-TH passive and active function sub-scale

Section A*	Measure	Infit	Outfit	Disordered threshold	Local dependence
Passive function sub-scale					
1. Cleaning Palm	0.08	0.93	0.89	Yes	No
2. Cutting finger nails	-0.79	0.79	0.77	Yes	No
3. Putting on a glove	-0.49	0.96	0.86	Yes	No
4. Cleaning armpit	-0.03	0.8	0.73	No	No
5. Putting arm through a sleeve	0.07	1.21	1.23	No	No
6. Put on a splint	1.06	1.31	0.89	Yes	No
7. Positioning arm on a cushion or support in sitting	0.09	1.19	1.31	Yes	No
Section B†					
Active function sub-scale					
1. Do up buttons on clothing	-0.39	0.96	0.87	Yes	No
2. Pick up a glass, bottle, or can	1.15	0.86	1.11	Yes	No
3. Use a key to unlock the door	-0.06	0.91	0.8	Yes	No
4. Write on paper	-0.54	1.22	1.22	Yes	No
5. Open a previously opened jar	-0.39	0.92	0.92	Yes	No
6. Eat with a knife and fork	0.55	0.94	0.8	Yes	No
7. Hold an object still while using unaffected hand	-1.66	1.38	1.75	Yes	No
8. Difficulty with balance when walking due to your arm	-0.87	1.12	1.27	Yes	No

*Asks about 'caring' for your affected arm either yourself with your unaffected arm or by a carer or a combination of both of these. This section does not ask about using your affected arm to complete any of the tasks. †Asks what you can do with your affected arm or using both arms.

Section A*	Measure	Infit	Outfit	Disordered threshold	Local dependence
Passive function sub-scale					
9. Dial a number on home phone	-0.2	1.07	0.88	Yes	No
10. Tuck in your shirt	-0.36	1.19	1.3	Yes	No
11. Comb or brush your hair	0.92	0.64	0.35	Yes	#12 #13
12. Brush your teeth	0.89	0.68	0.57	Yes	#13
13. Drink from a cup or mug	0.95	0.72	0.41	Yes	#12
*Asks about 'caring' for your affected arm either yourself with your unaffected arm or by a carer or a combination of both of these. This section does not ask about using your affected arm to complete any of the tasks. †Asks what you can do with your affected arm or using both arms.					

Reanalysis after rescoring from 5 to 3 response options; 0, 1 + 2, 3 + 4

The Eigen value of the first construct was reduced to 1.69 (13.5%). The disattenuated correlation between person measures was 1.00, and no local dependence was found, all suggest unidimensionality. The person reliability was reduced to 0.65. The Cronbach's alpha was 0.80. The item reliability was excellent (0.96). No disordered category threshold was found after this reanalysis.

For the ArmA-TH active function, all except item 7, fell within acceptable range of fit indices. This implied that item 7 could derail the Rasch measurement model (Table 2). Principal component analysis of residuals showed the 1st construct with an Eigen value of 2.57 (8.9%), suggesting violation of unidimensionality. The standardized residual correlations between item 13 and 12 was 0.59; item 11 and 12 was 0.52; and item 11 and item 13 was 0.45 indicating of local dependence as depicted and could be a source for another dimension (Table 2). However, the disattenuated correlation between person measures on the two item clusters was 1.00, suggesting that they were the same thing. The person reliability was acceptable (0.77), while the Cronbach's alpha was 0.85 and the item reliability was excellent (0.99). A disordered category was found in item 10 and disordered thresholds were found in all items. The ArmA-TH active function sub-scale did not appear to be well-targeted in this sample, as the mean logit between item and person was more than 1 SD. Item bias or DIF was not found in ArmA-TH active function on age and sex, and different education levels.

Reanalysis after rescoring from 5 to 3 response options; 0, 1 + 2, 3 + 4

The Eigen value of the 1st construct was reduced to 2.31 (8.2%). The standardized residual correlations between items 13 and 12 was 0.47; items 11 and 12 was 0.35; and items 11 and item 13 was 0.35, indicating some local dependence. However, the disattenuated correlation between person measures

was 1.00, suggesting sufficient unidimensionality. The person reliability increased to 0.78, Cronbach's alpha was 0.84 and the item reliability was excellent (0.98).

Transformation of raw scores to Rasch-scaled scores was illustrated in Table 3. Ideally, the ArmA raw score should be converted to the Rasch-scale score on the users' own data. However, this converted logit-scale should be applicable to situations where the data exhibit a similar fit to the model.

Table 3
Transformation of raw ArmA-TH scores to logits and then rescored to the original scale (n = 185)

ArmA-TH : Passive function					
Raw score	scale score	Raw score	scale score	Raw score	scale score
0	0	10	12	20	17
1	4	11	13	21	17
2	6	12	13	22	18
3	8	13	14	23	19
4	9	14	14	24	19
5	9	15	14	25	20
6	10	16	15	26	22
7	11	17	15	27	24
8	11	18	16	28	28
9	12	19	16		
ArmA-TH : Active function					
Raw score	scale score	Raw score	scale score	Raw score	scale score
0	0	18	22	36	28
1	7	19	22	37	28
2	10	20	23	38	29
3	12	21	23	39	29
4	14	22	23	40	30
5	15	23	24	41	30
6	16	24	24	42	31
7	16	25	24	43	32
8	17	26	24	44	32
9	18	27	25	45	33
10	18	28	25	46	34
11	19	29	25	47	35
12	19	30	26	48	37

ArmA-TH : Passive function						
13	20	31	26	49	39	
14	20	32	27	50	41	
15	21	33	27	51	45	
16	21	34	27	52	52	
17	21	35	28			

Discussion

This study aimed to explore measurement and scaling properties of the ArmA-TH using Rasch analysis in patients with hemiparetic upper limb. Our findings confirmed the unidimensionality for both the passive and the active function sub-scales. We found the same items with disordered threshold as did by Ashford et al (except for item 2). Although rescoring seemed to make the data fit the Rasch measurement model, it risks reducing person reliability in this passive sub-scale, which had fewer items compared to active sub-scale. However, some investigators had been less concerned by the disordered threshold, provided they do not impact on construct validity [8].

For the active function sub-scale, we found the original data did not fit well with Rasch measurement model, when compared to the passive function sub-scale. Four items were identified to be problematic. Item 7 “Hold an object still while using unaffected hand” did not contribute to the same construct compared to the other items. The high value of misfit indicated that this item was not productive, albeit, not harmful to the whole scale. Although items 11, 12, and 13 were dependent on each other to the extent that they could form a second dimension, the disattenuated correlation between person measure on the two item clusters suggested that they were the same thing.

Another point was that, all items in the active function sub-scale were found to have disordered thresholds, rescoring from 5 to 3 response options improved the fit with the Rasch model and was acceptable. The possible reason for the disordered thresholds might relate to limited comprehension of the rating scale by stroke patients due to cognitive impairment found in 20–80% of post stroke patients, present as early as 3–6 months after stroke onset [15, 16]. Category and threshold adjustment should be carefully considered for further exploration, particularly in the passive function sub-scale which had fewer items.

One limitation was that some participants required assistance to read the questionnaires because of vision or physical impairment. This may make their response influenced by the assistance, and interfere with their freedom in response.

Conclusions

According to results from Rasch analysis, both ArmA-TH active and passive function sub-scales data fit the Rasch model. Even though item 7 of the active function sub-scale seemed to present extra challenge to the Rasch model, the item was considered not harmful to overall measurement, provided useful clinical information and was therefore, retained. It is worth noting that a better fit to the model was observed when rescoring item responses from 5 to 3. Rescoring item response to less than 5 should be considered in future evaluation of ArmA-TH. Poor targeting in this sample implying that more easier items assessing arm function should be added.

Abbreviations

ArmA
Arm Activity Measure, ArmA-TH:Arm Activity Measure Thai version, CIT:Classic Test Theory, DIF:Differential Item Functioning, IRT:Item Response Theory, I-CVI:Content Validity Index for Item, S-CVI:Content Validity Index for Score.

Declarations

Ethics approval and consent to participate

Ethical approval for the research programme was received (number REH-2558-03109). All patients and caregivers (if present) were asked for written informed consent before proceeding with the questionnaire.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is supported by The Research Medical Fund, Grant Number 066/2559, Faculty of Medicine, Chiang Mai University.

Authors' contributions

MB, JK, PE, TW, SA participated in the conception and design of the study. MB, JK, PE, AK performed data collection. NW, TW performed the statistical analysis. TW, NW, MB, SA drafted and edited the manuscript. All authors made substantial contribution to the interpretation of data and revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

The authors thank all patients, their families and clinical colleagues who helped with this work. The authors thank personnel at Chiang Mai Neurological Hospital, Saraphi Hospital, the Northern Industrial Rehabilitation Center, and the research administration section, faculty of medicine Chiang Mai University for administrative assistance.

References

1. Ashford S, Slade M, Turner-Stokes L. Conceptualisation and development of the arm activity measure (ArmA) for assessment of activity in the hemiparetic arm. *Disabil Rehabil.* 2013;35:1513-8.
2. Ashford S, Turner-Stokes L, Siegert R, Slade M. Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm. *Clin Rehabil.* 2013;27:728-40.
3. Buntragulpontawee M, Euawongyarti P, Wongpakaran T, Ashford S, Rattanamanee S, Khunachiva J. Preliminary evaluation of the reliability, validity and feasibility of the arm activity measure - Thai version (ArmA-TH) in cerebrovascular patients with upper limb hemiplegia. *Health Qual Life Outcomes.* 2018;16:141.
4. Trevor G. Bond, Christine M. Fox. *Applying the Rasch model: fundamental measurement in the human sciences.* .New York: Routledge,Taylor & Francis group; 2015.
5. Ashford S, Siegert RJ, Alexandrescu R. Rasch measurement: the Arm Activity measure (ArmA) passive function sub-scale. *Disabil Rehabil.* 2016;38:384-90.
6. Linacre JM. *Winsteps® Rasch measurement computer program: User's Guide.* .Beaverton (OR): Winsteps.Com; 2017.
7. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical care.* 2004;42:I7-I16.
8. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007;57:1358-62.
9. Myers ND, Wolfe EW, Feltz DL, Penfield RD. Identifying differential item functioning of rating scale items with the Rasch model: An introduction and an application. *Meas Phys Educ Exerc Sci* 2006;10:215-40.

10. Institute for Objective Measurement, Inc. Validity and Rasch Measurement: Construct, Content, etc. <https://www.rasch.org/rmt/rmt181h.htm> (2005). Accessed 30 Apr 2020.
11. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas.* 2017;41:178-94.
12. Andrich. D. An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "Threshold disorder controversy". *Educ Psychol Meas.* 2013;73:78-124.
13. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch analysis of visual function and quality of life questionnaires. *Optom Vis Sci.* 2009;86:1160-8.
14. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas.* 2002;3:205-31.
15. Madureira S, Guerreiro M, Ferro JM. Dementia and cognitive impairment three months after stroke. *Eur J Neurol.* 2001;8:621-7.
16. Sun J-H, Tan L, Yu J-T. Post-stroke cognitive impairment: epidemiology, mechanisms and management. *Ann Transl Med.* 2014;2:80.