

Disability Measurement for Korean Community-Dwelling Adults With Stroke: Item-Level Psychometric Analysis of the Korean Longitudinal Study of Ageing

Ickpyo Hong, PhD, OTR¹, Young Joo Kim, PhD, OTR/L², Mandi L. Sonnenfeld, MOT, OTR³, Emily Grattan, PhD, OTR/L⁴, Timothy A. Reistetter, PhD, OTR¹

¹Department of Occupational Therapy, School of Health Professions, University of Texas Medical Branch, Galveston, TX; ²Department of Occupational Therapy, College of Allied Health Science, East Carolina University, Greenville, NC; ³Division of Rehabilitation Sciences, School of Health Professions, University of Texas Medical Branch, Galveston, TX; ⁴Department of Health Sciences and Research, College of Health Professions, Medical University of South Carolina, Charleston, SC, USA

Objective To investigate the psychometric properties of the activities of daily living (ADL) instrument used in the analysis of Korean Longitudinal Study of Ageing (KLoSA) dataset.

Methods A retrospective study was carried out involving 2006 KLoSA records of community-dwelling adults diagnosed with stroke. The ADL instrument used for the analysis of KLoSA included 17 items, which were analyzed using Rasch modeling to develop a robust outcome measure. The unidimensionality of the ADL instrument was examined based on confirmatory factor analysis with a one-factor model. Item-level psychometric analysis of the ADL instrument included fit statistics, internal consistency, precision, and the item difficulty hierarchy.

Results The study sample included a total of 201 community-dwelling adults (1.5% of the Korean population with an age over 45 years; mean age=70.0 years, SD=9.7) having a history of stroke. The ADL instrument demonstrated unidimensional construct. Two misfit items, money management (mean square [MnSq]=1.56, standardized Z-statistics [ZSTD]=2.3) and phone use (MnSq=1.78, ZSTD=2.3) were removed from the analysis. The remaining 15 items demonstrated good item fit, high internal consistency (person reliability=0.91), and good precision (person strata=3.48). The instrument precisely estimated person measures within a wide range of theta (-4.75 logits < θ < 3.97 logits) and a reliability of 0.9, with a conceptual hierarchy of item difficulty.

Conclusion The findings indicate that the 15 ADL items met Rasch expectations of unidimensionality and demonstrated good psychometric properties. It is proposed that the validated ADL instrument can be used as a primary outcome measure for assessing longitudinal disability trajectories in the Korean adult population and can be employed for comparative analysis of international disability across national aging studies.

Keywords Stroke, Aging, Community survey, Reliability and validity, Outcome assessment

Received May 16, 2017; Accepted July 27, 2017

Corresponding author: Ickpyo Hong

Department of Occupational Therapy, School of Health Professions, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1142, USA. Tel: +1-409-772-3082, Fax: +1-409-747-1615, E-mail: ichong@utmb.edu

ORCID: Ickpyo Hong (<https://orcid.org/0000-0001-5404-7646>); Young Joo Kim (<https://orcid.org/0000-0002-6258-7463>); Mandi L. Sonnenfeld (<https://orcid.org/0000-0001-6958-2641>); Emily Grattan (<https://orcid.org/0000-0002-2714-4130>); Timothy A. Reistetter (<https://orcid.org/0000-0003-1732-5533>).

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited. Copyright © 2018 by Korean Academy of Rehabilitation Medicine

INTRODUCTION

Since the 1990s, Western and European countries have launched longitudinal national surveys targeting the rapidly aging population [1]. The purpose of these national surveys is to obtain a better understanding of aging at the population level and establish social and health policies. The Korean Longitudinal Study of Ageing (KLoSA) was implemented in 2006 and data of nationally representative Korean adults have been collected every even-numbered year.

The Korean Labor Institute designed the KLoSA to comprehensively assess health aspects of the rapidly growing Korean aging population, including family structure, health, activities of daily living (ADL), psychological aspects of aging, employment, income, and subjective expectations. In particular, the KLoSA aimed to create a longitudinal Korean aging survey for cross-cultural comparisons comparable with various international aging surveys [1-3]. In contrast to other aging studies, the KLoSA included a young adult population aged between 45 and 49 years because of the devastating financial crisis in the late 1990s, which resulted in the massive dismissal of middle-aged Korean adults [4].

Although the majority of research has studied stroke among the elderly, recent epidemiologic studies by Korean researchers have studied the association between the younger population and stroke [5]. Generally, stroke has a negative effect on not only physical function and quality of life but also imposes a substantial economic burden on the individual as well as on the country [6]. Over 100,000 Korean adults experience a new or recurrent stroke every year, with a stroke incident occurring on average every 5 minutes. The annual stroke incidence in Korea is estimated to increase to 350,000 by 2030 [5,7]. Due to the increase in the elderly population and the fact that the average age of stroke onset is getting younger, it is imperative we understand the impact of stroke on functional status within the population.

Since a number of aging studies have been implemented, the US National Institute on Aging has focused on harmonizing international population-based longitudinal aging studies to enable researchers and health policy makers to conduct cross-national comparisons across various international aging populations [8]. While the survey contents of these national aging surveys have

been harmonized, international health comparisons are still not feasible due to incomparable outcome measures across different countries. Traditionally, disability is estimated by the functional status of ADLs, such as eating, dressing, grooming, toileting, etc. [9-11], and disability severity has played a critical role in establishing health policies and services for the aging population [12]. Although the harmonized national surveys contain a set of ADL items [1-4], the number of items (17 ADLs in KLoSA and 10 ADLs in the Health and Retirement Study [HRS]) and their responses categories (3-point rating scale in KLoSA and dichotomous rating scale in HRS) were different in the surveys, which hindered direct disability comparisons across various countries [1].

The measurement challenge is a well-known limitation of classical test theory which uses total scores of instruments to compare outcomes [13,14], and it is not often feasible to compare instruments that have different measurement constructs (number of items and rating scales). To address this measurement challenge, a 1-parameter model (i.e., Rasch model) of item response theory (IRT) has been applied to the harmonized national surveys [9,10,15]. The reported research methodology is highly applicable to the KLoSA since one of the primary purposes of the KLoSA is to construct an international comparative study on the aging population [4]. However, until now, no attempts have been made to calibrate the ADL items of the KLoSA using the Rasch model.

Therefore, the primary objective of this study was to use the Rasch model for the creation of an internationally uniform ADL disability metric for community-dwelling aging adults with stroke living in Korea. We hypothesized the ADL items of the KLoSA as unidimensional, so it would be feasible to calibrate them by the Rasch model. We also hypothesized that a Rasch-calibrated KLoSA measure would allow researchers to conduct international disability comparison studies with respect to other national aging studies, which have been previously calibrated by the Rasch model. In addition, the calibrated ADL items could be used as a stable disability outcome measure for assessing longitudinal disability trajectories among the Korean aging population.

MATERIALS AND METHODS

Design and participants

We obtained de-identified study data from the first wave of the 2006 KLoSA containing participants' data for Korean community-dwelling adults aged 45 years or over [4]. As per the KLoSA study protocol, trained surveyors collected informed consents from participants and conducted face-to-face interviews using a computer-assisted personal interviewing program. The inclusion criteria for the current study sample were: (1) having a stroke diagnosis by a physician during the past 12 months and (2) having difficulties in performing ADL items. Due to the potential ceiling effect on disability measurement typically observed among the healthy population, we excluded participants who had no difficulty performing ADL items [16]. As the KLoSA is a national, public-opened data with de-identification, this study was exempted by the Institutional Review Board of the University of Texas Medical Branch (No. 17-0127).

Outcome measures

Activities of daily living

The KLoSA adapted existing basic ADL (BADL) and instrumental ADL (IADL) instruments to assess the functional status of the community-dwelling adult population [17-19]. The KLoSA consists of 7 BADL items, including dressing, washing the face, bathing, eating, getting out of bed, toileting, and bladder/bowel management and 10 IADL items, including grooming, housekeeping, preparing meals, laundering, going out, using public transportation, shopping, money management, phone use, and medication management. The 17 ADL items were coded using a 3-point rating scale: 1 (no assistance needed), 3 (partial assistance needed), and 5 (total assistance needed), with a possible score ranging from 17 to 85 points. To calibrate the ADL items, we changed the rating scale values to 1, 2, and 3 so that the total scores ranged from 17 to 51 points. We also reversed the rating scales of the ADL items so that higher scores intuitively reflect higher independence status on daily tasks.

Data analysis

The psychometric properties of the ADL instrument were tested using IRT methodologies recommended by the patient-reported outcome measure information

system (PROMIS) groups [20]. First, the unidimensionality assumption was examined using confirmatory factor analysis (CFA). Upon meeting the unidimensionality assumption, further Rasch analysis was applied to examine the psychometric properties of the ADL instrument at the item level, including (1) fit statistics, (2) precision statistics (internal consistency, person strata, and standard error curve), and (3) item difficulty hierarchy statistics [14].

Unidimensionality

The unidimensionality of the ADL instrument was investigated by performing CFA with a one-factor model. We used the weighted least squares with adjustments for the mean and variance (WLSMV) estimation for the categorical response categories of the ADL instrument [20,21]. The unidimensionality of the ADL instrument was analyzed using model fit indices, including chi-square (χ^2)/degrees of freedom (df) test (<3.84 indicating $p>0.05$), comparative fit index (CFI) >0.95 , Tucker-Lewis Index (TLI) >0.95 , and root mean square error of approximation (RMSEA) <0.08 [20]. Examination of the amount of variance explained by the dominant factor structure was conducted by calculating the Omega coefficient ($\hat{\rho}_{X,\omega}$) with the standardized loadings of each item on the dominant factor structure. Factor loadings greater than 0.70 were considered adequate [22]. We also examined local independence, one of the IRT core assumptions, using the residual correlation coefficient extracted from the dominant factor. We considered test items having a residual correlation value greater than or equal to 0.2 as problematic items [20,21].

Rasch analysis

Rasch analysis was used to examine the psychometric properties of the identified measurement structure(s) using the Andrich rating-scale model (RSM) [23]. First, the internal validity of the test items was analyzed using infit (information-weighted) and outfit (un-weighted) mean square (MnSq) statistics that have approximate chi-square distributions [24,25]. Fit statistics were used to examine if individuals responded to test items in an expected manner estimated by the Rasch measurement model [25,26]. MnSq indicates the amount of distortion in the measurement with a reported ideal value of 1.0 with a range from 0 to positive infinity [27]. Likewise, the acceptable range of infit and outfit MnSq values for surveys is 0.60 to 1.40

with an acceptable range for the standardized z-statistics (ZSTD) of -2.00 to 2.00 [27,28]. Finally, if test items demonstrated misfit in the Rasch model, we removed them from the ADL instrument and recalibrated the instrument with non-misfit items only. This process was iterated until we obtained a final set of items that exhibited sufficient fit with the Rasch model [29].

The internal consistency of the ADL instrument was analyzed using the Rasch person reliability index, which provides a comparative analysis similar to the traditional Cronbach’s alpha statistic. Person reliability over 0.80 is considered a satisfactory value [30,31]. The precision of the ADL instrument was presented as person strata which are defined as distinct measurement groups that are centered three measurement errors apart from one another [32]. Person strata were calculated using the following formula:

$$\text{Person strata} = \frac{(\text{person separation index} \times 4 + 1)}{3}$$

At least three person strata are considered acceptable and equivalent to a reliability of 0.8 [30,31]. Standard error (SE) values across the spectrum of the latent trait (-5.0 logits to 5.0 logits) were also analyzed. The SE cutoff was calculated using the following formula:

$$SE = S \times \sqrt{(1-r)}$$

where S is the standard deviation of person measures estimated by the Rasch model and r is a pre-specified measurement reliability of 0.9 [33].

The unique strength of the Rasch model relies on the fact that person ability and item difficulty are calibrated into the same linear interval scale with logits scores [14,26]. Based on a linear scale, the item difficulty hierarchy and the item-person match were investigated using the Rasch person-item map. We considered 1.0 logit difference between the mean of the item difficulty and person measure as an acceptable match [34]. For clinical use, a scoring table is provided which converts the theta values into T scores with an average of 50 and a standard deviation of 10.

We used Mplus version 7.4 software to perform CFA and to conduct calculations of the omega coefficient as well as residual correlations [21]. WINSTEPS Rasch software version 3.92.2 was used for Rasch analysis [24]. SAS version 9.4

software (SAS Institute Inc., Cary, NC, USA) was used to create an analytical file and conduct descriptive statistics.

Sample size calculation

Mathematically, the RSM requires a sample of 64 to 144

Table 1. Demographics of the study sample (n=201)

Variable	Value
Age (yr)	70.0±9.7
Sex	
Male	118 (58.1)
Female	83 (41.3)
Education attainment	
Elementary school	134 (66.6)
Middle school	24 (11.9)
High school	33 (16.4)
Beyond college	10 (4.9)
Living area	
Metropolitan	83 (41.2)
Suburban	68 (33.8)
Rural	50 (24.8)
Living alone	48 (23.8)
Regular exercise	51 (25.4)
Current smoker	28 (13.9)
Current drinker	29 (14.4)
Current employment	5 (2.5)
BMI (kg/m ²)	22.5±3.0
Number of assistive devices	0.8±0.7
Household income (10,000 KRW/yr)	1,094±1,654
Chronic conditions ^{a)}	
Hypertension	106 (52.7)
Depression	67 (33.3)
Diabetes	47 (23.3)
Arthritis	38 (18.9)
Hearing problem	37 (18.4)
Psychopathy	34 (16.9)
Cataract	26 (12.9)
Cardiac disorders	23 (11.4)
Cancer	9 (4.5)
Lung disorders	7 (3.5)
Liver disorders	5 (2.5)

Values are presented as mean±standard deviation or number (%).

BMI, body mass index; KRW, Korean won.

^{a)}All participants have multiple chronic conditions.

observations to obtain stable item calibrations (± 0.5 logit at a 95% confidence interval) and to assure that the calibrations of all items are more than 0.5 logit away from the Rasch model calibration [35]. Also, Wang and Chen [36] reported that the RSM resulted in negligible item bias estimations with as few as 100 subjects using Monte Carlo simulation techniques.

RESULTS

The database contained data of 10,218 participants including 361 community-dwelling adults with stroke. These individuals accounted for 2.97% (weighted sample size=469,927) of the Korean adult population in 2006 who were 45 years or older. Two hundred one of these community-dwelling adults with stroke (weighted sample size=251,303, 1.5% of the Korean population over 45 years old) demonstrated difficulty performing ADLs and were included in the study. Table 1 presents the demographic characteristics of the study sample. The majority of the sample included males ($n=118$, 58.1%), individuals who had completed elementary school ($n=134$, 66.6%), and individuals with hypertension ($n=106$, 52.7%).

Unidimensionality

The ADL instrument reflected a unidimensional mea-

surement structure by demonstrating good model fit statistics to a CFA one-factor model: $\chi^2/df=2.99$, CFI=0.994, TLI=0.993, and RMSEA=0.100. All factor loadings on the dominant factor were greater than 0.888 ($0.888 < \lambda < 0.987$) and the ADL instrument explained 94.4% variance ($\hat{\rho}_{X,\omega} = 0.944$) in observed summative scores across the 17 ADL items. These findings indicate that the ADL instrument met the unidimensionality assumption and measures a single construct. There was no violation of the local independence assumption, and all the residual correlations ranged from 0.000 to 0.118 (< 0.20). Since these findings supported a unidimensional ADL instrument, Rasch analysis was subsequently performed.

Rasch analysis

The initial Rasch item calibration demonstrated the money management item (infit MnSq=1.31, ZSTD=2.5; outfit MnSq=1.56, ZSTD=2.3) and phone use item (infit MnSq=1.44, ZSTD=3.0; outfit MnSq=1.78, ZSTD=2.3) misfit the Rasch model. These two items were removed and re-calibration was conducted for the remaining 15 items since misfit items can affect the person measure estimations. Table 2 presents item fit statistics and item difficulty hierarchy of the ADL instrument for the final 15 items. Once the two misfit items were removed, the remaining items sufficiently met the Rasch model as-

Table 2. Item fit and item difficulty hierarchy results

Item	Measure (logits)	Model SE	Infit		Outfit	
			MnSq	ZSTD	MnSq	ZSTD
Laundering	2.57	0.16	1.01	0.1	0.85	-0.4
Using public transportation	2.44	0.16	1.17	1.4	1.03	0.2
Preparing meals	1.89	0.16	0.93	-0.6	0.95	-0.1
Housekeeping	1.35	0.16	1.01	0.1	0.93	-0.2
Going out	1.09	0.16	1.12	1.0	0.96	-0.1
Shopping	0.93	0.16	1.20	1.6	1.04	0.3
Bathing	0.51	0.17	0.64	-3.3	0.55	-2.4
Dressing	-0.47	0.18	0.99	-0.1	1.23	0.9
Grooming	-0.60	0.19	1.01	0.1	1.33	1.1
Washing the face	-0.64	0.19	0.94	-0.4	0.86	-0.4
Getting out of bed	-1.36	0.20	1.08	0.6	0.63	-0.8
Medication management	-1.36	0.20	1.05	0.4	0.67	-0.7
Eating	-1.84	0.22	1.33	1.9	1.37	0.8
Toileting	-2.04	0.22	0.79	-1.3	0.45	-0.9
Bladder/bowel management	-2.47	0.24	1.07	0.5	0.58	-0.6

SE, standard error; MnSq, mean square; ZSTD, standardized z-statistics.

sumptions. The Rasch model estimated a mean person measure logit of 1.49 (SD=2.47) with a mean of test item logit of 0.00 (SD=1.60).

The ADL instrument demonstrated a high person reliability of 0.91 and high person strata of 3.48, indicating that the ADL instrument can separate the study sample into approximately 3 measurement groups. Fig. 1 presents the precision of the ADL instrument across the latent trait with a reliability of 0.90. The cut-off of SE (dash line) was estimated by the standard deviation of the person measure (2.47) and a pre-specified reliability of 0.9. The ADL instrument precisely estimated person measures located in a wide range of theta from -4.75 logits to 3.97 logits at a reliability level of 0.9.

Fig. 2 presents the person-item map. Person measures are located on the left-side of the map. On the right-side, test items are located with a median 50% cumulative probability between the adjacent rating scales (Rasch half-point step thresholds). The person-item map indicates that the ADL items sufficiently covered the person measure distribution.

The most difficult item was the laundering item (2.57 logits) and the easiest item was the bladder/bowel management item (-2.47 logits). The # marks at the bottom indicate people who had the lowest person ability and the # marks at the top indicate people who had highest person ability. While the rating scale of the ADL instrument covered a wide range of theta, there was a floor effect (n=36,

17.9%) since there were individuals who needed total assistance across all of the 15 items. In addition, for the clinical use, the Rasch logits scale was transformed into a mean of 50 and standard deviation of 10 (Table 3).

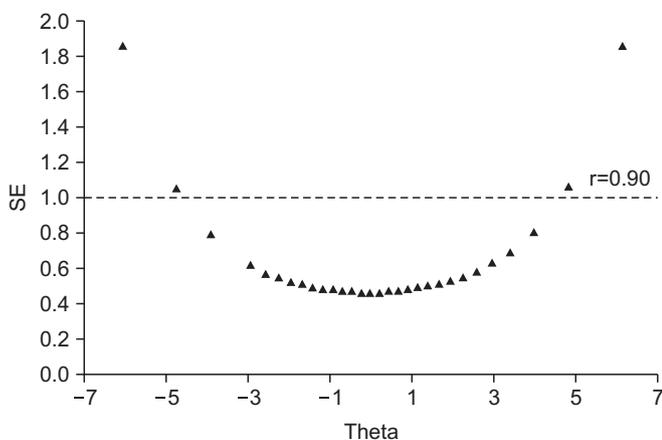


Fig. 1. Standard error (SE) curve across the latent trait (θ). Dash line indicates the cut-off for SE which is equivalent to a reliability of 0.9. The activities of daily living instrument precisely estimated person measures located in a wide range of theta from -4.75 logits to 3.97 logits at a reliability level of 0.9.

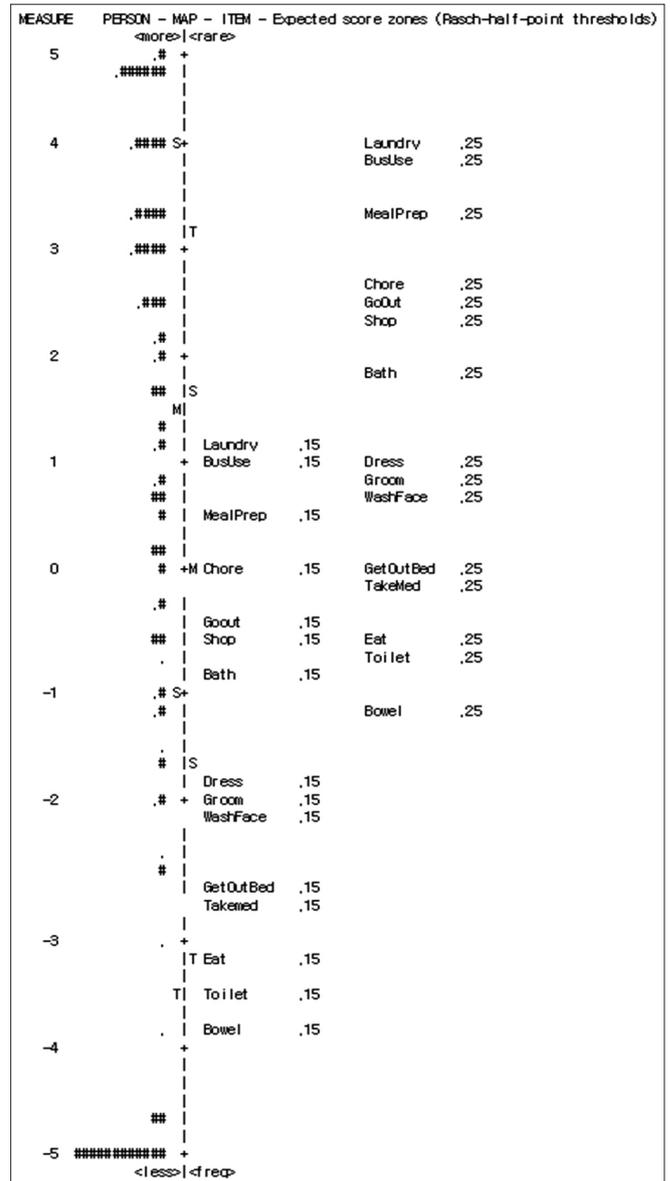


Fig. 2. Person item-map. The numbers on the left side indicate the measures of person ability and item difficulty in logits. Each # on the left side of the map indicates three subjects and items are located on the right side of the map. The M to the left of the vertical line is the mean of the person measures and M to the right is the mean of the item difficulties. 'Ss' on the left and right of the vertical line indicates 1 standard deviation and the 'Ts' on the left and right of the vertical line indicates 2 standard deviations for item difficulty and person ability, respectively.

Table 3. Disability scale with the T scores conversion of the raw score into interval-level scores on a mean of 50 and standard deviation of 10

Total score (15-45)	Measure	Rescaled interval score	SE
15 ^{a)}	-6.05	-10.52	18.63
16	-4.75	2.48	10.67
17	-3.92	10.81	8.00
18	-3.37	16.27	6.89
19	-2.94	20.56	6.26
20	-2.58	24.22	5.85
21	-2.25	27.46	5.56
22	-1.96	30.43	5.34
23	-1.68	33.19	5.18
24	-1.42	35.80	5.04
25	-1.17	38.29	4.94
26	-0.93	40.69	4.86
27	-0.70	43.03	4.80
28	-0.47	45.32	4.76
29	-0.24	47.57	4.74
30	-0.02	49.82	4.74
31	0.21	52.07	4.75
32	0.43	54.33	4.78
33	0.66	56.64	4.82
34	0.90	59.00	4.89
35	1.14	61.43	4.98
36	1.40	63.96	5.09
37	1.66	66.63	5.23
38	1.95	69.46	5.41
39	2.25	72.51	5.65
40	2.59	75.87	5.95
41	2.96	79.65	6.37
42	3.41	84.10	7.01
43	3.97	89.73	8.11
44	4.82	98.24	10.76
45 ^{a)}	6.14	111.38	18.68

SE, standard error.

^{a)}The measure estimations at the extremes are not accurate due to the high measurement errors of the Rasch model.

DISCUSSION

The findings of the present study revealed that 15 ADL items of the 2006 KLoSA data met the unidimensional assumption of the IRT model supporting the use of these

items for measuring disability severity among community-dwelling adults with a history of stroke. The validated ADL instrument can be used with other cross-national aging studies that have been harmonized using Rasch modeling to compare disability. In addition, the calibrated ADL items can be used as a robust baseline measure for examining longitudinal disability trajectories among the Korean adult population 45 years or older.

The study findings demonstrated superior psychometric properties compared to the Survey of Health, Ageing and Retirement in Europe (SHARE) which was also calibrated by a Rasch model [2,9]. While the KLoSA demonstrated high precision (person reliability=0.91, person strata=3.48), the SHARE demonstrated moderate precision (person reliability=0.74, person strata=2.66). In addition, the KLoSA ADL instrument precisely estimated person measures located in a wide range of theta (-4.75 logits to 3.97 logits) compared to the SHARE ADL Instrument (-0.74 logits to -0.13 logits) at a reliability of 0.9. These differences in the precision may be due to population and survey variation across the 16 European countries included in the SHARE data. In contrast, the KLoSA was implemented in Korea only, and the well-trained single country surveyors may have reduced administration errors and measurement variation. As the KLoSA ADL instrument demonstrated high precision, it is expected to be used as a primary outcome measure for longitudinal disability studies in the Korean population across different measurement time periods (2006, 2008, 2010, 2012, and 2014 KLoSA data).

The ADL instrument revealed two misfit items including the money management (infit MnSq=1.31, ZSTD=2.5; outfit MnSq=1.56, ZSTD=2.3) and the phone use (infit MnSq=1.44, ZSTD=3.0; outfit MnSq=1.78, ZSTD=2.3); therefore, these items were removed to secure the invariance of the ADL instrument [29] because misfit items could have led to distortion of the item calibration among the 15 non-misfit items. We speculate that these two items may have misfit because they are more cognitively demanding. For instance, in contrast to the majority of the other BADL and IADL items, these two items require higher-level cognitive (e.g., calculation, executive function) or language functions. These two items are also less physically demanding than most of the other IADLs (e.g., meal preparation). Another explanation would be that these items are frequently performed by caregivers

or someone else because these IADLs are not as critical to disability as other IADLs like medication management which is considered to be an urgent safety task requiring cognitive function. In addition, money management and phone use are often activities that are either performed by one individual in the household or are shared activities performed by multiple members within a household. Lastly, the misfit for these two items may have been due to the narrow scope of the sample population, which were community-dwelling adults with a mean age of 70 years. Since the item calibrations and fit statistics were based on a specific population in this study, future studies will need to examine the item fit of the two items across various chronic conditions and age groups

The ADL items demonstrated a conceptual item difficulty hierarchy. The most difficult items were the laundering (2.57 logits) and using public transportation (2.44 logits), and the easiest items were the bladder/bowel management (-2.47 logits) and toileting (-2.04 logits). Traditionally, IADL items are considered more challenging than BADL items, and similar item difficulty patterns have been reported in the HRS, SHARE, and the English Longitudinal Study of Ageing (ELSA) in the United Kingdom [3,9,10]. The results of the present study reveal that KLoSA has similar construct validity when compared to other Rasch calibrated disability measurements used in aging studies, indicating that the KLoSA ADL instrument should strongly be considered for use in cross-national disability studies.

Although the four national aging studies, KLoSA, HRS, SHARE, and ELSA, have been calibrated by the same IRT approach (Rasch model) and have demonstrated comparable psychometric properties, cross-national measurement challenges still exist. One challenge is that cross national instruments use different metrics to develop disability measures. For example, the HRS and ELSA were co-calibrated using ADLs as well as physical (grip strength and balance) and cognitive items (depression and memory) [10], whereas the SHARE included ADLs with mobility items (walking 100 m, climbing one flight of stairs without resting) [9], without any specific cognitive items. A possible solution to address these measurement differences is to use common test items across Rasch-calibrated aging studies with a Rasch common-item equating method [8]. In the Rasch common-item equating method, a set of ADL items, across aging studies are used as anchor points

for equating surveys. Using these common items, a co-calibrated measure can be developed which allows for disability comparisons across different countries.

Our study has several limitations. First, the Rasch-calibrated KLoSA items were not validated by well-accepted ADL instruments, such as the Functional Independence Measure or Barth Index. Future studies are warranted to validate the Rasch-calibrated items against standardized ADL instruments. Second, the KLoSA ADL instrument was based on data from community-dwelling adults with stroke, which may not generalize to other disability populations. In addition, we limited our target population to those with mild to moderate disability because this group has a considerable impact on healthcare utilization and cost. Additionally, typical adults without impairments have an influence on disability measurement due to high ceiling effects. For example, a previous study demonstrated high ADL ceiling effects among the general population of community-dwelling adults (young adults or fully independent in ADLs) [16]. For these reasons, our study was limited to individuals with a history of stroke whose functional status was lower than the general population.

In conclusion, to the best of our knowledge, this is the first study to calibrate KLoSA data and examine the psychometric properties of the ADL instrument at an item-level using a Rasch model. The study findings indicate that the Rasch-calibrated ADL instrument can be used as a robust outcome measure for assessing the severity of disability in Korean community-dwelling adults with stroke. Considering its excellent psychometric properties, the Rasch-calibrated ADL instrument is expected to be utilized for cross-national disability comparisons as well as to serve as a precise metric for future longitudinal disability trajectory studies in community-dwelling Korean adults with stroke.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

We thank the Korea Employment Information Service for providing the study data (<http://survey.keis.or.kr/eng/index.jsp>).

REFERENCES

1. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol* 2014;43:576-85.
2. Borsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, et al. Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol* 2013;42:992-1001.
3. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. *Int J Epidemiol* 2013;42:1640-8.
4. Boo KC, Chang JY. Korean longitudinal study of ageing: research design for international comparative studies. *Surv Res* 2006;7:97-122.
5. Hong KS, Bang OY, Kang DW, Yu KH, Bae HJ, Lee JS, et al. Stroke statistics in Korea: part I. Epidemiology and risk factors: a report from the Korean stroke society and clinical research center for stroke. *J Stroke* 2013;15:2-20.
6. Lim SJ, Kim HJ, Nam CM, Chang HS, Jang YH, Kim S, et al. Socioeconomic costs of stroke in Korea: estimated from the Korea national health insurance claims database. *J Prev Med Public Health* 2009;42:251-60.
7. Hong KS, Bang OY, Kim JS, Heo JH, Yu KH, Bae HJ, et al. Stroke statistics in Korea: part II stroke awareness and acute stroke care, a report from the Korean Stroke Society and Clinical Research Center for Stroke. *J Stroke* 2013;15:67-77.
8. Shih RA, Lee J, Das L. Harmonization of cross-national studies of aging to the health and retirement study: cognition. Santa Monica: RAND Corporation; 2012.
9. Buz J, Cortes-Rodriguez M. Measurement of the severity of disability in community-dwelling adults and older adults: interval-level measures for accurate comparisons in large survey data sets. *BMJ Open* 2016;6:e011842.
10. Cieza A, Oberhauser C, Bickenbach J, Jones RN, Ustun TB, Kostanjsek N, et al. The English are healthier than the Americans: really? *Int J Epidemiol* 2015;44:229-38.
11. Grimby G, Andren E, Holmgren E, Wright B, Linacre JM, Sundh V. Structure of a combination of Functional Independence Measure and Instrumental Activity Measure items in community-living persons: a study of individuals with cerebral palsy and spina bifida. *Arch Phys Med Rehabil* 1996;77:1109-14.
12. Altman BM. Definitions, concepts, and measures of disability. *Ann Epidemiol* 2014;24:2-7.
13. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989;70:857-60.
14. Wright BD, Stone MH. Best test design. Chicago, IL: Mesa Press; 1979.
15. Chan KS, Kasper JD, Brandt J, Pezzin LE. Measurement equivalence in ADL and IADL difficulty across international surveys of aging: findings from the HRS, SHARE, and ELSA. *J Gerontol B Psychol Sci Soc Sci* 2012;67:121-32.
16. Hong I, Yoo EY, Kazley AS, Lee D, Li CY, Reistetter TA. Development and validation of the activities of daily living short form for community-dwelling Korean stroke survivors. *Eval Health Prof* 2018;41:44-66.
17. Won CW, Rho YG, Kim SY, Cho BR, Lee YS. The validity and reliability of Korean Activities of Daily Living(K-ADL) Scale. *J Korean Geriatr Soc* 2002;6:98-106.
18. Won CW, Rho YG, SunWoo D, Lee YS. The validity and reliability of Korean Instrumental Activities of Daily Living(K-IADL) Scale. *J Korean Geriatr Soc* 2002;6:273-80.
19. Won CW, Yang KY, Rho YG, Kim SY, Lee EJ, Yoon JL, et al. The development of Korean Activities of Daily Living(K-ADL) and Korean Instrumental Activities of Daily Living(K-IADL) Scale. *J Korean Geriatr Soc* 2002;6:107-20.
20. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45(5 Suppl 1):S22-31.
21. Muthen LK, Muthen BO. Mplus version 7.4 software [Internet]. Los Angeles: statmodel.com; 2015 [cited 2018 Mar 1]. Available from: <https://www.statmodel.com/index.shtml>.
22. Tabachnick BG, Fidell LS. Using multivariate statistics. 5th ed. Boston: Pearson; 2006.
23. Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978;43:561-73.
24. Linacre JM. A user's guide to WINSTEPS v3.93.2 [Internet]. [place unknown]: WINSTEPS; 2017 [cited 2018 Mar 1]. Available from: <http://www.winsteps.com/winman/copyright.htm>.

25. Wright BD. Despair and hope for educational measurement. *Contemp Educ Rev* 1984;3:281-8.
26. Bond TG, Fox CM. *Applying the Rasch model fundamental measurement in the human sciences*. Mahwah: Psychology Press; 2001.
27. Linacre JM. What do infit and outfit, mean-square and standardization mean? *Rasch Meas Trans* 2002;16:878.
28. Wright BD, Linacre JM, Gustafson JE, Martin-Lof P. Reasonable mean-square fit values. *Rasch Meas Trans* 1994;8:370.
29. Linacre JM. When to stop removing items and persons in Rasch misfit analysis. *Rasch Meas Trans* 2010;23:1241.
30. Fisher WP. Reliability, separation, strata statistics. *Rasch Meas Trans* 1992;6:238.
31. Fisher W. Rating scale instrument quality criteria. *Rasch Meas Trans* 2007;21:1095.
32. Wright BD, Masters GN. *Rating scale analysis*. Chicago: Mesa Press; 1982.
33. Harvill LM. Standard error of measurement. *Educ Meas* 1991;10:33-41.
34. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care* 2004;42(1 Suppl):I17-24.
35. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans* 1994;7:328.
36. Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the WIN-STEPS program for the family of Rasch models. *Educ Psychol Meas* 2005;65:376-404.