A Rasch analysis of the Sunnaas ADL Index for stroke survivors

Av Kristin S. Loxley & Ann-Kristin G. Elvrum

Kristin S. Loxley is an occupational therapist, MSc, working at Sunnaas Rehabilitation Hospital and affiliated with Department of Interdisciplinary Health Sciences, Faculty of Medicine, University of Oslo, Norway.
Email: kristin.loxley@sunnaas.no

Ann-Kristin G. Elvrum is an occupational therapist, ph.d., working at Clinic of Rehabilitation, St. Olav's Hospital, Trondheim University Hospital, Trondheim and Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim.

Abstract

Sunnaas ADL Index's internal scale validity

Background: The Sunnaas ADL Index (SADL) is a generic outcome measure that evaluates independence in activities of daily living (ADL). Acceptable reliability of SADL for patients with stroke has been confirmed, while information concerning construct validity is lacking.

Aims/Objectives: The objective of our study was to evaluate SADL's internal scale validity and aspects of reliability for stroke survivors.

Material and Methods: Data from 200 patients (136 men, mean age 58 years) admitted to the stroke unit at Sunnaas Rehabilitation Hospital (2012-2017), was included in the Rasch measurement model analyses to evaluate internal scale validity and item and person reliability of SADL.

Results: SADL was found to be unidimensional with an item fit of 91.8 percent and excellent person (0.91) and item reliability (0.99). Only one item, communication, showed misfit, but was retained, as removal did not improve fit statistics. The scoring categories functioned well, but mean person ability was higher than mean item difficulty with nine persons reaching maximum scores.

Conclusion: The SADL is a valid measure of independence in daily activities for stroke survivors and seems most appropriate for those with moderate to severe disabilities. It should be considered to revise the scoring criteria for the item communication.

Keywords: Outcome Measures; Model, Statistical; Psychometrics; Activities of Daily Living

Introduction

Loss of independence in activities of daily living (ADLs) is a devastating consequence for many persons who experience stroke and other brain related injuries (1, 2). It affects their possibility to live independently and safely at home and to participate in work and leisure activities. Regaining independence in daily activities is associated with higher quality of life, reduced caregiver burden and less use of healthcare resources (3, 4). Occupational Therapists (OTs) are especially concerned with enabling performance and participation in everyday activities (5). Therefore, assessments of activity performance are essential for OT's working with rehabilitation of people with acquired brain injuries. Evaluating the performance and level of independence in both primary self-care activities (PADLs) such as eating and dressing and instrumental activities (IADLs) such as cooking and shopping, provide a starting point for goal-setting and planning of interventions and adaptations to promote independence in ADLs (1). To evaluate individual treatment effect, as well as measuring the level of independence, safety, required level of assistance and potential for rehabilitation, assessments of ADL with sounds psychometric properties are crucial (3, 6). In addition, valid and reliable outcome measures that are sensitive to change, are important to provide evidence-based knowledge concerning efficacy of various interventions (1, 3, 7).

The Sunnaas ADL Index (SADL) is commonly used by occupational therapists in Norway to evaluate independence in ADLs (8, 9). In addition, the

Figure 1. Scoring sheet for the Sunnaas ADL Index.

Barthel Index and Functional Independence Measure (FIM) are frequently used to measure ADL outcomes (3). The SADL assesses the level of independence in completing twelve daily activities as well as how much assistance and/or adaptation is needed to complete them (10). Unlike Barthel Index or FIM, the SADL includes items evaluating instrumental ADLs that are relevant for OT interventions (cooking, housework and outdoor mobility) (11). The scoring sheet also provides a visual overview over the level of assistance and adap-

tations needed to perform the included activities (Figure 1) (12). This enables the therapist to easily illustrate for the patient what steps are needed to become independent in these activities and discuss possible interventions.

Although originally created for patients with stroke, the SADL has also been used for geriatric patients and patients with conditions such as spinal cord injuries, multiple trauma, poliomyelitis and osteogenesis imperfecta (8, 10, 13-18). Earlier studies have investigated the psychometric properties of SADL for the joint groups

of stroke and geriatric patients as well as for patients with various neurological diagnosis (14-16, 19, 20). Those studies have indicated that the SADL has good reliability, whereas the findings regarding the scale's validity were less consistent (15, 19-22). The latter specifically concerns whether each item on the SADL assesses the same underlying construct and the functioning of the intervals of the scoring scale (14, 15). Hence, there is still a need to investigate further the internal construct validity of the SADL scale for the group of stroke patients.

Rasch measurement model analysis is the recommended method of evaluating the internal validity and reliability of the item scale for criterion-referenced outcome measures used in rehabilitation (23). Criterion-referenced outcome measures evaluate the patient's performance against pre-specified criteria describing different levels of efficient or independent performance (24). The resulting raw scores are ordinal and do not exhibit basic features of measurement, such as unidimensionality, hierarchical order and equal interval scaling. The use of Rasch measurement model analysis transforms ordinal data into interval equal level data, enabling the use of parametric statistical analyses. In addition, the internal scale validity and reliability of the outcome measure can be established, for example, whether the construct measured by the test items (for example independence in ADLs) remains stable over the range of person abilities in the population of interest (25). Thus far, only one study has used Rasch analysis to investigate the construct validity of the SADL (14). That study, inclu-

Characteristics	
Participants (n (%))	200 (100)
Gender (n (%))	
Men	136 (68)
Women	64 (32)
Age in years (Mean (±SD))	58 (12)
Men	56 (12.6)
Women	59 (11.7)
Diagnosis (n (%))	
Stroke	188 (94)
Others*	12 (6)
T.S.O. (Median (min-max))	33 (6-3195)
*Brain tumour, encephalitis, other neurological diagnosis, n= Number of participants, SD=Standard Deviation; min= minimum, max= maximum, T.S.O. = Time since onset given in days	

Table I. Characteristics of included persons with stroke and other acquired brain injuries.

ding a heterogeneous sample of geriatric patients, indicated that several items did not measure the same construct as the other items (14). However, methodological weaknesses, including missing descriptions of statistical analyses and the resulting estimates, make it difficult to draw firm conclusions from the study. Furthermore, the results for a geriatric patient group may not apply to the population of patients with stroke, for whom the SADL was originally developed.

The aim of our study was therefore to examine the construct validity of the SADL for patients with stroke by exploring its internal scale validity and aspects of its reliability (internal consistency, Rasch analysis based item- and person reliability coefficients and person separation ratios).

Methods

STUDY DESIGN AND SAMPLE
We used convenience sampling in our cross-sectional study. The

inclusion criteria were being a patient of at least 18 years old with stroke or similar motor and cognitive impairments following an acquired brain injury. The SADL was administered to 200 patients admitted to the stroke unit at Sunnaas Rehabilitation Hospital in Norway from 2012 to 2017 (see Table I). The stroke unit is a secondary care unit with patients that are referred for complex rehabilitation from primary hospitals in the South-East Health Region of Norway. The average age and length of stay for patients admitted to the stroke unit at Sunnaas Rehabilitation Hospital in 2017 was 54,8 years and 46.5 days (26). The number of persons included in the study was chosen according to recommendations from Linacre (27) in order to obtain precise, robust measurements.

THE SUNNAAS ADL INDEX (SADL)
The SADL was developed in 1985 by Norwegian occupational the-

rapists at Sunnaas Rehabilitation Hospital (10). The assessment tool was intended to provide an overview over the areas of ADL's relevant to consider the patient's independence. It contains twelve ADL items:

1 eating
2 continence
3 indoor mobility
4 toilet management
5 transfer
6 dressing and undressing
7 grooming
8 cooking
9 bathing and/or showering
10 housework
11 outdoor mobility
12 communication.

Each item has four ordinal scoring categories ranging from 0 to 3 points, for a maximum sum score of 36 points. Scores from 0 to 1 indicate total or partial dependence on assistance, whereas scores of 2 to 3 indicate independence with or without adaptation or assistive devices. The items are structured into three hierarchical groups based on frequency of activity performance and degree of assistance needed. Items 1 to 4 (eating, continence, indoor mobility and toilet management) refer to activities performed several times daily that cannot be planned to occur at a specific time. By contrast, items 5 to 8 (transfer, dressing and undressing, grooming and cooking) refer to activities also performed several times daily, but that can be arranged to occur at a specific time. Items 9 to 11 (bathing and/or showering, housework and outdoor mobility) refer to activities that can be planned, but occur only once or twice a week. Last, item 12, communication, is not included in any group due to its frequency. The SADL should be administered and scored according to specific criteria outlined in the manual (10).

DATA COLLECTION
The SADL was administered as part of regular clinical evaluations by occupational therapists or occupational therapy students who were trained in using the SADL. The assessment was completed as an interview with the patient and/or relatives and supplemented by clinical observations when considered to be necessary.

ETHICS
The study was approved as a Quality Improvement Project by the Data Protection Officer at Oslo University Hospital, Norway. The approved aim of the study was to investigate the measurement properties of the SADL, not to investigate characteristics of the individuals in the group or to produce new knowledge about people or disease. Thus, the study was not required to be evaluated by the Regional Committee for Medical and Health Research Ethics in Norway, and consent was not obtained from the participants (28). Data was de-identified with the link key stored separately. Thereafter, the data was handled and analysed pseudonymously.

DATA ANALYSIS
The Rasch measurement model was used to evaluate the SADL's internal scale validity and aspects of item and person reliability using Winsteps version 3.71.0.1. Six areas were investigated in an iterative analytical process:
(i)   functioning of the rating scale
(ii) unidimensionality
(iii) targeting of item difficulty to person ability
(iv) item and person reliability
(v)   item invariance
(vi) hierarchical structure

The rating scale model and the functioning of the rating scale
First, rating scale functioning was investigated to determine which derivation of the Rasch polytomous model to use for further analysis (29). For the rating scale to function well, the threshold values should increase by >1.4 logits between each category. A minimum of ten responses for each category is recommended for the scale to function as expected (30).

Unidimensionality
Unidimensionality was investigated by principal components analysis (PCA) and item and person goodness-of-fit statistics. The PCA may indicate existence of a secondary dimension if the eigenvalue in the first contrast amounts to more than two (31). We also investigated potential multidimensionality by exploring if there was clustering between groups of items. This is more important than whether the loadings exceed certain values (31). Local independence of the items was also explored. Since local dependency may inflate reliability indices, analyses were repeated with testlets (correlated items pooled together) if the standardized residual correlation between two items exceeded 0.3 (32).

Goodness-of-fit statistics indicate how well the items fit the underlying construct and how close to the expected value the persons perform (30). Since misfitting infit statistics pose a greater threat to test validity than misfitting outfit statistics, criterion for an acceptable infit mean square (MnSq)

Nr	Item Name	Measure	Model S.E.	Infit MnSq	Infit Zstd	Outfit MnSq	Outfit Zstd
10	Housework	2.57	0.14	0.93	-0.6	0.94	-0.2
11	Outdoor Mobility	1.49	0.12	0,98	-0.2	0.89	-0.8
8	Cooking	1.42	0.12	1.18	1.6	1.14	1.0
9	Bath/shower	0.82	0.12	0.70	-3.2	0.60	-3.7
3	Indoor Mobility	0.21	0.12	0.82	-1.7	0.80	-1.5
6	Dressing	-0.07	0.12	0.92	-0.6	0.98	0.0
4	Toilet Management	-0.12	0.12	0.47	-5.8	0.43	-4.8
5	Transfer	-0.18	0.12	0.42	-6.6	0.38	-5.4
7	Grooming	-0.44	0.12	0.50	-5.3	0.54	-3.3
1	Eating	-1.61	0.12	1.33	2.7	2.51	4.2
2	Continence	-1.62	0.12	1.31	2.5	1.02	0.2
12	Communication	-2.47	0.13	2.17	7.2	9.90	9.9
SADL = Sunnaas ADL Index, ADL = Activities of daily living, Nr = item number, S.E. = Standard Error, MnSq = Mean Square of residuals, Zstd= standardized t-values.							

Table II. Overview of misfit among SADL items, with items presented in hierarchical order from most to least difficult, numbers in bold indicate item misfit.

was set to infit MnSq >1.5 in combination with a standardised Z value (Zstd) >2.0 for the items (33). Mean square values exceeding 2.0 can be degrading for the measurement tool. With those criteria, we could identify items showing underfit in relation to the Rasch model (33). In the iterative process, misfitting items and persons were removed, after which the analyses were repeated.

Targeting of item difficulty to person ability
When exploring item–person targeting, an equally wide range in item difficulties and person abilities is desired, with an even distribution of persons and items across the entire continuum of the scale (30). To examine targeting, the estimated mean person- and item-related measures and standard deviations (SD) were reviewed. The mean person ability measure indicates how difficult the items are for the persons in the sample, whereas the SD indicates the range of person abilities and item difficulties. An item–person map was also investigated, which illustrates how the range of person abilities combines with the range of item difficulties (30).

Item and person reliability of the rating scale
We investigated the item and person reliability coefficients, including Cronbach's alpha and person separation ratios (30). Item reliability indicates whether the hierarchical item estimates can be replicated across other samples, whereas person reliability captures the ability of the scale to reliably rank estimated person-related measures. It is recommended that values exceed 0.9 for item reliability and 0.8 for person reliability. Lower values could indicate too few persons in the sample or that the instrument is not sensitive enough to differentiate high from low performers (34). The person separation value (G) was used to calculate the number of ability levels that the scale can differentiate using the formula (4G + 1) / 3 (35). The standard errors indicated the precision of the measurements (30).

Item invariance
Testing for differential item functioning (DIF) was used to investigate invariance of the SADL scale (30). Invariance requires that the items function the same way independently of different attributes of the sample being measured such as gender and age. The DIF contrast should be at least 0.5 logits with a statistical probability of p<0.05 to be noticeable (36). For age, the groups were divided by the median of 59 years, with Group 1 consisting of persons aged ≤59 years and Group 2 of persons aged ≥60 years.

Hierarchical structure
The Rasch analysis calculates a hierarchical structure based on the average difficulty of each item. The hierarchy of item difficulty was examined and compared with SADL's intended hierarchical group structure (10, 11).

Results

SAMPLE
The mean age of participants was 58 years (SD: 12 years) (Table I), which was substantially lower than the mean age of patients with stroke in Norway in 2017 (females = 77 years, males = 72 years) (37). The gender distribution, with 68 percent men, was higher than reported for patients

with stroke on a national level (54 percent men), but was in accordance with the gender distribution of patients admitted to the stroke unit at Sunnaas Rehabilitation Hospital in 2017 (68 percent men) (26, 38).

RATING SCALE FUNCTIONING
Rasch analysis of rating scale functioning indicated that the SADL scale functioned in accordance with the expectations using the rating scale model, meaning that the 4-point rating scale had at least ten observations for each category when all items were considered, and the Rasch–Andrich thresholds increased in value as the category number increased (29). However, the difference in threshold logits between Category 2 (.98) and Category 3 (1.34) was less than the recommended range of 1.4. Category 2 also had the least number of observations (n = 361). Compared to the number of observations in categories 0, 1 and 3, a clear underuse of category 2 was found. In iterative analyses, Category 2 was collapsed with Category 3 without creating any mentionable changes to the fit statistics. As a result, we decided to keep all four categories, because Category 2 was considered as clinically relevant to document changes in ADL function and to provide essential patient information at discharge planning.

UNIDIMENSIONALITY
The results of the PCA showed that 73.9 percent of the variance in the data was explained by the SADL measures, which indicated that the data fit the model well. The eigenvalue of the unexplained variance in the first contrast amo-
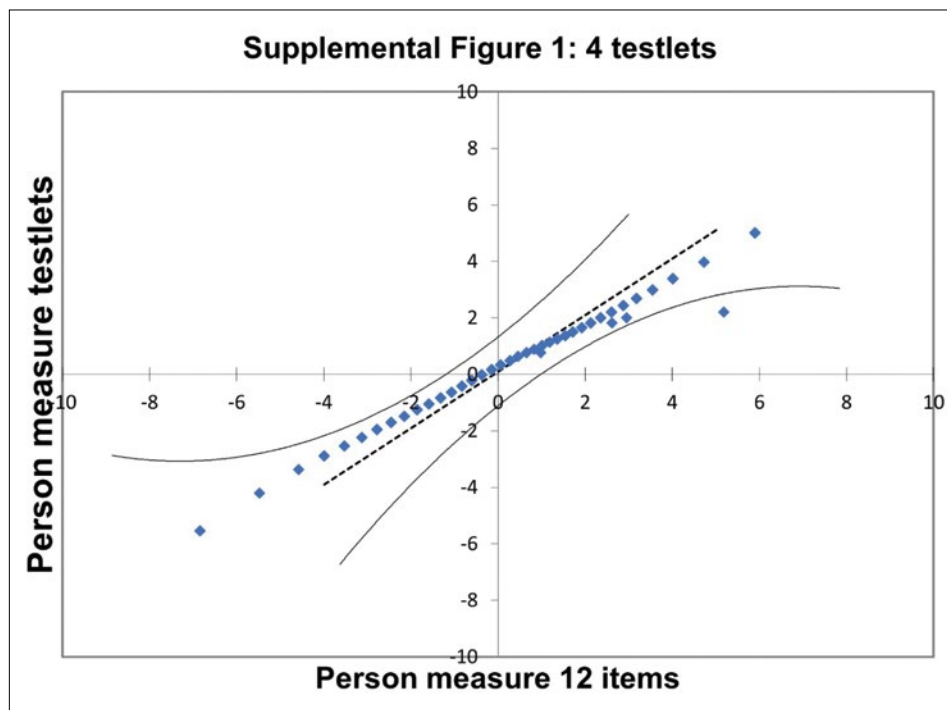
Supplemental Figure 1.

unted to 2.4 (5.2 percent). When examining the plot over item loadings, we found no obvious clustering, except between items toilet management and transfer with a local dependency correlation of 0.59. Another item pair with potential local dependency was dressing and grooming (0.34). We therefore combined the two item pairs with local dependency into testlets and ran new analyses. This reduced the eigenvalue of the first contrast to 2.1 and created a minimal change in person reliability from 0.92 to 0.91, while item reliability was constant at 0.99. When cross-plotting the person measures from the analysis including all twelve items against those from the analysis including testlets, the measures were well aligned indicating no noticeable differences (Supplemental Figure 1). This result, together with no obvious clustering in the first contrast, indicated that the SADL in the practical and

clinical sense could be regarded as unidimensional (39, 40).

Infit statistics also indicated unidimensionality with 11 of 12 items fitting the scale (91.8 percent item fit). Item 12 (communication) was the only misfitting item (infit MnSq = -2.47, Zstd = 2.17), as shown in Table II. This result indicated that communication measured a different construct than the other items. However, removing the item did not improve item fit, because other items became more misfitting. Within the Rasch model, a 95 percent item fit is acceptable; therefore, we decided to retain the item in the scale (30). Items 4 and 5 (toilet management and transfer) were overfitting, with MnSq values <0.5. Because overfitting items do not degrade the scale and both items were considered clinically important, those two items were also retained. Item 12 (communication) also displayed large misfitting
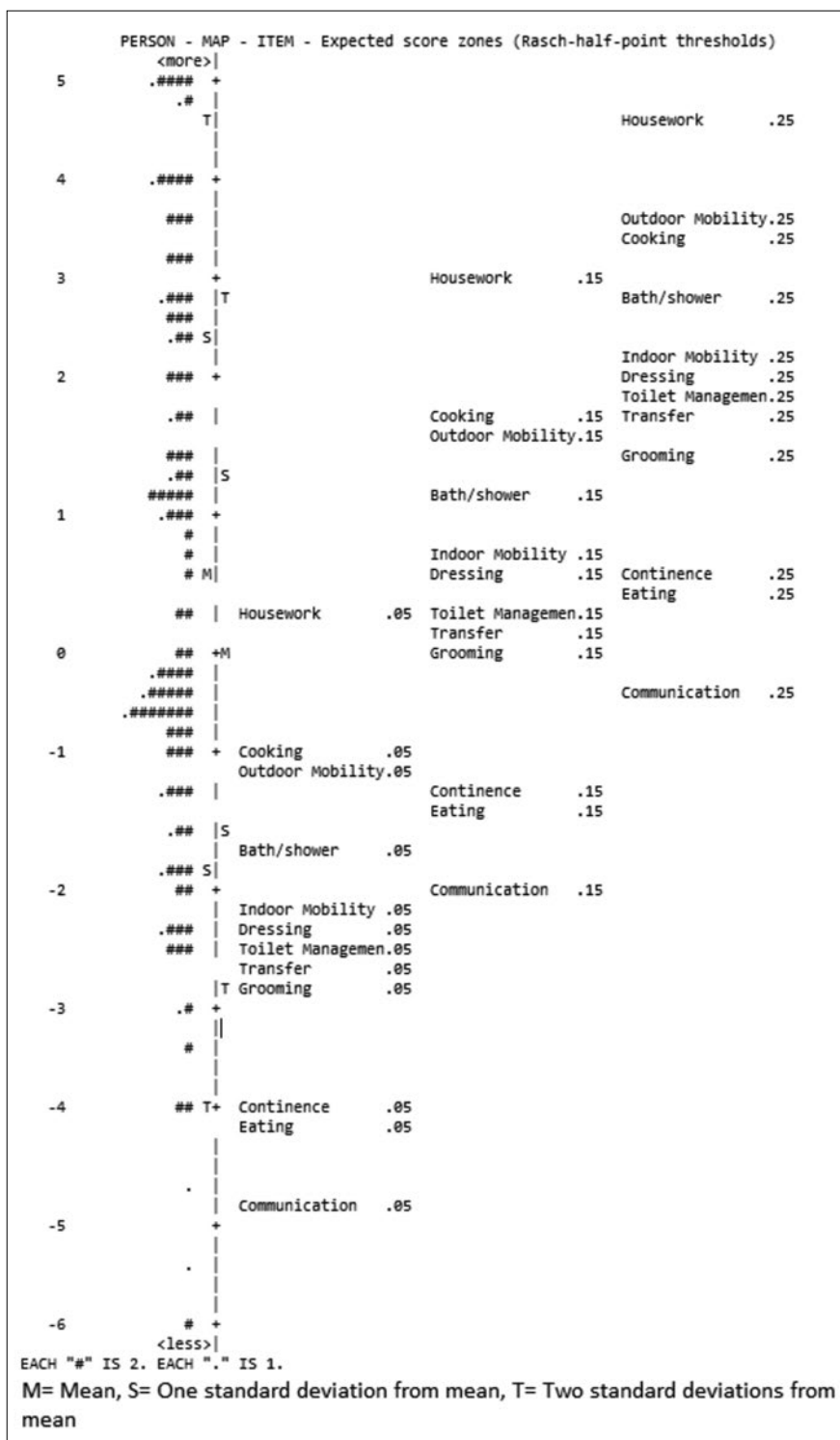
```
          PERSON - MAP - ITEM - Expected score zones (Rasch-half-point thresholds)
                 <more>|
     5      .####  +
              .#   |
                  T|                                      Housework       .25
                   |
                   |
     4      .####   +
                   |
              ###  |                                      Outdoor Mobility.25
                   |                                      Cooking          .25
              ###  |
     3             +                   Housework    .15
             .###  |T                                     Bath/shower      .25
              ###  |
              .## S|
                   |                                      Indoor Mobility .25
     2        ###  +                                      Dressing         .25
                   |                                      Toilet Managemen.25
              .##  |           Cooking         .15        Transfer         .25
                   |           Outdoor Mobility.15
              ###  |
              .## |S                                      Grooming         .25
            #####  |           Bath/shower     .15
     1       .###  +
               #   |
               #   |           Indoor Mobility .15
               #  M|           Dressing        .15        Continence       .25
                   |                                      Eating           .25
              ##   |  Housework        .05  Toilet Managemen.15
     0        ##  +M            Transfer         .15
            .####  |            Grooming         .15
            .####  |
          .#######  |                                      Communication    .25
              ###  |
    -1       ###  +  Cooking          .05
                   |   Outdoor Mobility.05
             .###  |           Continence       .15
                   |           Eating           .15
              .## |S
                   |  Bath/shower      .05
             .### S|
    -2        ##  +            Communication    .15
                   |  Indoor Mobility .05
             .###  |  Dressing        .05
              ###  |  Toilet Managemen.05
                   |  Transfer         .05
                  |T Grooming          .05
    -3        .#  +
               #  ||
                   |
                   |
    -4        ## T+  Continence       .05
                   |  Eating           .05
                   |
                   |
               .   |
    -5             +  Communication    .05
                   |
                   |
               .   |
                   |
    -6         #  +
                 <less>|
EACH "#" IS 2. EACH "." IS 1.
M= Mean, S= One standard deviation from mean, T= Two standard deviations from
mean
```

*Figure 2. Item-person map of distribution of person abilities and item difficulty with
expected scores between scoring categories.*

outfit statistics (MnSq = 9.9, Zstd
= 9.9), which indicated that some
persons scored higher or lower
on the item than their estimated
ability.

**TARGETING OF ITEM DIFFICULTY
TO PERSON ABILITY**
Altogether, nine persons achieved
maximum scores, whereas two
persons had minimum scores,
which constituted a 4.5 percent
ceiling and 1 percent floor effect.
Together with a mean person-abi-
lity measure of 0.51, those results
indicate a slight overweight of
persons with high abilities in the
sample (30). Both the standard
deviations of items and persons
(Table III) and the item–person
map (Figure 2) indicated that the
variation among person abilities
was wider than among the item
difficulties. The item–person map
also illustrated that several items
were in the middle of the range
for difficulty (indoor mobility,
dressing and undressing, transfer,
toilet management and groo-
ming), whereas fewer items had
high difficulty.

**ASPECTS OF THE ITEM AND
PERSON RELIABILITY OF THE
RATING SCALE**
Reliability and separation values
for both persons and items were
within the desired range (Table
III) (39). Cronbach's alpha was
.95, which suggests excellent
internal consistency. The standard
errors of items were reasonably
small, which indicates relatively
precise measurement of the item
difficulties, whereas the larger
standard error of the estimated
person ability measures indica-
ted more imprecision (Table III).
Calculations based on the person
separation value showed that the
sample of participants could be
differentiated into 4.7 different
ability levels.

ITEM INVARIANCE
DIF analysis indicated that item
12 (communication) functioned
differently for persons of high
versus low age, with a statistically
significant DIF-contrast of 0.84
logits (p=0.02). No statistically
significant DIF-contrast was found

Area of analysis	Statistical approach, criteria	Results SADL	Conclusion
The rating scale functioning (whether the rating scale meets the expectations of the Rasch model).	The threshold values between categories should increase by >1.4 logits and a minimum of 10 observations for each category is required.	The threshold values increased between categories, but threshold between category 2 and 3 was smaller than recommended. Category 2 had the least number of observations (n = 361).	Results indicate acceptable rating scale functioning. No categories should be removed due to clinical importance, consider revising criterion for category 2.
Principal components analysis (PCA) to explore unidimensionality (whether the items measure the same construct, or if several dimensions exist)	Loading <2 in first contrast could indicate multidimensionality, with additional clustering between groups of items. Residual correlations of <0.3 indicate local dependency and may inflate reliability indices. Correlated item pairs can be combined into testlets to explore impact on reliability.	73.9 % of the data variance was explained by SADL with an eigenvalue of 2.4 in the first contrast, and no obvious clustering. Two item pairs had local dependency (correlations: 0.59 and 0.34) and were combined into testlets. New analysis with testlets showed that reliability indices remained high (person reliability changed from 0.92 to 0.91 and item reliability remained at 0.99)	Results indicate that SADL measure one main construct supporting unidimensionality. Local dependency of two item pairs did not cause noteworthy inflation of reliability indices.
Item goodness-of-fit statistics (how well the item responses match the expected responses of the Rasch Model)	Criterion for Item fit: Infit Mean Square (MnSq) <1.5 in combination with standardised Z value (Zstd) <2.0. Acceptable with 5 % item misfit.	11 of 12 items were within acceptable criterion for item fit (91.8 % of the items in the scale). Item 12 (communication) was the only misfitting item (infit MnSq = -2.47, Zstd = 2.17).	Results indicate acceptable item fit and SADL unidimensionality. Removing item 12 (communication) did not improve item fit. Consider revising this item for future research.
Targeting of item difficulty to person ability (how well the range of item difficulty functions with the range of person abilities)	An equally wide range in item difficulties and person abilities, with an even distribution of persons and items across the entire continuum of the scale is desired.	9 persons with maximum scores, 2 persons with minimum scores (4.5 % ceiling and 1 % floor effect). Standard Deviations of items and persons and the item–person map indicated a slightly wider variation among person abilities than among the item difficulties.	Results indicate acceptable targeting. The ceiling effect indicates that SADL might be more suitable for persons with moderate to severe disability.
Item and person reliability of the rating scale (can the scale reliably rank item difficulty and person ability measures? What is the precision of the instrument?)	Recommended values; Person reliability: >0.9 Item reliability: >0.8 Cronbach's alpha: >0.7 Higher S.E. indicate less precision. The person separation value: used to calculate the number of ability levels that the scale can differentiate between.	Reliability values for both persons (0.92) and items (0.99) were above the desired values. Cronbach's alpha (.95) suggests excellent internal consistency. Item S.E. were reasonably small (0.14) and indicates precise measurement of item difficulties, a larger S.E. of person ability measures (0.72) indicated more imprecision. Calculations based on the person separation value (G=3.30) indicate that participants can be differentiated into 4.7 ability levels.	The SADL shows very good reliability values, indicating precision of the item scale hierarchy. Further, the results indicate that the scale can differentiate between almost 5 ability levels of independence in ADLs.
Item invariance (whether the items function differently for different groups of people)	The Differential Item Functioning (DIF) contrast should be at least 0.5 logits with a statistical probability of p<0.05 to be noticeable. This was investigated for age and gender.	Item 12 (communication) functioned differently for persons of high versus low age, with a statistically significant DIF-contrast of 0.84 logits (p=0.02). The DIF-contrast was not statistically significant for any items for gender.	Results indicate that younger persons in this sample were more independent in communication than older persons.
Hierarchical structure: (How is the hierarchy of item difficulty compared with SADL's intended hierarchical group structure?)	The Rasch analysis calculates a hierarchical structure based on the average difficulty of each item.	Three of the four most difficult items (housework, outdoor mobility and bath/shower) were in the group of activities performed once or twice weekly. Furthermore, the two least difficult items (eating and continence) belonged to the group of activities performed daily that could not be preplanned, whereas indoor mobility was found to be a more difficult item.	The hierarchical order of the item difficulty measures partly conformed to the intended group structure, indicating that most patients with stroke are independent in eating and continence, while housework is far more difficult.

n = number of persons, SD = Standard deviation, S.E. = Standard errors

Table III. Overview of areas of rating scale analysis with the Rasch measurement model for the Sunnaas ADL Index (SADL).

for gender. This indicates that younger persons in this sample were more independent in communication than older persons.

HIERARCHICAL STRUCTURE
The hierarchical order of the item difficulty measures partly conformed to the intended group structure. The IADL items for housework and outdoor mobility were the most difficult, whereas communication was the least difficult item according to the Rasch analyses (Table II). The items with the closest item difficulty were toilet management and transfer. Three of the four most difficult items (housework, outdoor mobility and bath/shower) were in the group of activities performed once or twice weekly (Table II). Furthermore, the two least difficult items (eating and continence) belonged to the group of activities performed daily that could not be preplanned, whereas indoor mobility was found to be a more difficult item. Three of four activities performed daily that could be preplanned (transfer, dressing and undressing and grooming) had a similar order of difficulty, whereas the last, cooking, was the third most difficult item.

Discussion
The results provide evidence of the SADL's internal scale validity, as well as item and person reliability, as a measure of independence in ADLs for patients with stroke. The Rasch analyses indicated that the SADL was unidimensional with an item fit of 91.8 percent. Only one item, communication, did not fit the model. Communication was the only item with potential DIF, as well as the least difficult of the SADL items, while housework

was the most difficult item. Some local dependence was identified and corrected utilising testlets, without significantly impacting on the results on reliability. Lastly, the Rasch analyses indicated that the SADL can differentiate between nearly five ability levels among stroke survivors. This finding may imply that the SADL is a measurement tool that can be sensitive to change, although that possibility needs to be further explored.

Our study is the first to demonstrate internal scale validity of the SADL for patients with stroke. The measurement scale, including the measurement steps, functioned as expected. Furthermore, the eigenvalue and the infit statistics indicated acceptable unidimensionality (30). This unidimensionality is noteworthy given the relatively complex construct that frame independence in ADL. In accordance with our results, Vardeberg (15) found that the SADL consisted of a single dimension, whereas Nielsen & Wæhrens (14) did not find evidence of scale validity due to several misfitting items and poorly ordered measurement steps. These contrasting results may be caused by differences in the patient samples. The study by Vardeberg (15) addressed various neurological diagnoses, including stroke, whereas Nielsen & Wæhrens (14) included a geriatric patient group. This may indicate that items included in the SADL are better fitting for post-stroke patients than for geriatric patients.

The only misfitting item in our analyses was communication. That result aligns with the findings in previous studies (14, 15). The misfit might be caused by unexpected low communication scores for which some persons

with estimated high abilities, or vice versa. Speech impairments such as aphasia do not necessarily affect independence in ADL (41), and communication might therefore be the most difficult item for physically able patients with aphasia, despite being the least difficult item for most of the patients. Another reason for the misfit could be that the complexity of communication is not reflected in the SADL manual, such that patients can achieve the maximum score for communication despite being able to communicate their basic needs only (42). The ability to communicate verbally and physically following stroke can be influenced by several factors, including motor paralysis of face and throat muscles, as well as cognitive impairments (43). A Rasch analysis performed on the Functional Independence Measure also indicated that communication measured a dimension other than the motor-based items (44).

In validating the SADL, we decided to keep the communication item despite its misfit, because removing it only led to the misfit of other items. ADL outcome measures such as the SADL consist of various items, or daily activities, that are required to reflect independence in ADL. Removing too many items may result in an underrepresentation of the construct being assessed and a serious loss of information. Thus, it is recommended to use clinical judgement in addition to considerations of item fit in the validation of an outcome measures (45). Because communication disorders can affect a stroke survivor's required level of assistance in IADLs (46), we considered it important to retain that item in the SADL. However, for future

studies, modifying the description of communication in the SADL manual should be considered to improve item fit and perhaps also remove potential DIF between patients at younger and older ages.

Our results also confirm the reliability of the SADL by indicating trustworthiness in its internal consistency, the reproducibility of the item hierarchy and the distribution of person abilities. Some SADL items displayed local dependency, and the person reliability was slightly lowered from 0.92 to 0.91 when using testlets to absorb the effect of local dependency. The item pairs with local dependence appear to share some similar traits, for example transfer to or from a wheelchair for the items transfer and toilet management. However, combining these two items would mean that important information regarding the person's independence in ADL would be lost. Thus, the content of the items is crucial to consider when handling local dependency (31, 47). The findings regarding internal scale reliability are in accordance with the results of previous studies (13, 15). Past studies have also revealed strong correlations between SADL and other ADL outcome measures, thereby indicating the satisfactory concurrent validity of the SADL (13, 19).

Overall, the hierarchy of difficulty of the SADL items established by the Rasch analysis reflects clinical judgement of item complexity, and we found some similarities with the intended group structure of the SADL. The three most difficult items according to the Rasch analysis (housework, indoor mobility and cooking) are all instrumen-

tal activities. IADLs are more complex activities that require the higher functioning of physical and cognitive capacities than P-ADLs (1). Two of those activities (housework and outdoor mobility) belong to the SADL group of activities performed once or twice weekly, whereas cooking needs to be performed every day. Thus, the information provided in the hierarchy of item difficulty (Table II) may be useful, in addition to the SADL intended group structure, concerning the level of assistance that a person would need to be able to live at home.

Results describing the targeting of the SADL show that the range of item difficulties is somewhat narrower than the range of person abilities (illustrated in Figure 2). In our sample, 4.5 percent of persons reached the ceiling, which indicates that the SADL was too easy for some patients and may not have enough items to assess in detail potential small changes in ADLs for patients with mild disabilities caused by stroke. However, for most patients with stroke, the SADL seems to function well and will probably add useful information in treatment planning and at discharge when considering the need for assistance and adaptations required for independence in ADL.

Furthermore, our results indicate that the SADL can differentiate between nearly five different ability levels among stroke patients. That finding may imply that the SADL is sensitive to change, as previously suggested by Korpelainen, Niilekselä (19), who found the SADL to be responsive and able to detect clinically important changes in P-ADLs and I-ADLs for stroke survivors. Another study, however, found

that the SADL could not identify the minimal detectable change in a geriatric population (13). Those conflicting results, possibly due to different patient populations, also indicate the need for further research to gain more robust knowledge about the responsiveness of the SADL.

LIMITATIONS
Our study has some limitations. First, our sample included patients undergoing rehabilitation in a secondary care unit, had more men and a lower mean age than in the population of patients with stroke on a national level (37). However, analyses indicated that the items overall did not function differently between men and women or between older and younger persons. Only the communication item functioned differently between younger and older age groups. Thus, it is unlikely that this influenced our main result indicating evidence of the SADL's internal scale validity. Second, data in the study was collected as a part of the usual clinical evaluations. Thus, variation might have occurred in how different therapists asked questions and scored the SADL. However, previous studies have indicated the SADL's good reliability (13, 15, 16), and the use of data collected in the clinic may be regarded as an advantage for the clinical relevance of our results. Third, our patient sample does not include patients with stroke alone, but also patients with other brain-related injuries with similar physical and cognitive impairments. This could have had an impact on the generalization of our results. However, we found no larger changes in the results when excluding these persons from the analyses, which

may indicate similar impairments as the persons with stroke. Last, large variation in time since onset may have contributed to the wide range of person abilities in our study; however, this variation may also be representative of the large group of patients with stroke. Thus, we do not think that the variation systematically affected our results.

Conclusion

Our study indicates acceptable internal scale validity of the Sunnaas ADL Index for stroke survivors, which suggests that the SADL can provide valid information about their level of independence in performing ADLs. Only one item, communication, did not fit the model, and revision of the scoring criteria for that item is therefore recommended. The SADL scale also displayed good reliability, and the rating scale categories functioned as expected according to the Rasch model. A limited number of patients with stroke reached the ceiling of the SADL. Thus, the SADL can be recommended to validly measure independence in ADL's in patients with stroke overall. However, if the aim is to plan treatment and consider the need of assistance or adaptation in ADL's, then the SADL in probably more useful for those with moderate to severe disabilities. Further studies are required to investigate the SADL's validity in other patient groups as well as its responsiveness.

Acknowledgements

References

1. Tuntland H. En innføring i ADL : teori og intervensjon. Kristiansand: Høyskoleforlaget; 2006.
2. Teasell R, Hussein N. Background Concepts in Stroke Rehabilitation. 2018 March 2018. In: Evidence-Based Review of Stroke Rehabilitation [Internet]. ebrsr.com: Heart and Stroke Foundation - Canadian Partnership for Stroke Recovery; [34]. Available from: http://www.ebrsr.com/sites/default/files/v18-SREBR-CH3-NET.pdf.
3. van Der Putten JJMF, Hobart JC, Freeman JA, Thompson AJ. Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. Journal of Neurology, Neurosurgery and Psychiatry. 1999;66(4):480-4.
4. de Weerd L, Rutgers WAF, Groenier KH, van der Meer K. Perceived wellbeing of patients one year post stroke in general practice - recommendations for quality aftercare. BMC Neurology. 2011;11(1):42.
5. Ergoterapeutene. Alle skal kunne delta - Ergoterapeutenes kjernekompetanse. ergoterapeutene.org/ergoterapi: Norsk Ergoterapeutforbund; 2017. p. 32.
6. Salter K, Campbell N, Richardson M, Mehta S, Jutai J, Zettler L, et al. Outcome measures in stroke rehabilitation. Evidence-Based Review of Stroke Rehabilitation [Internet]. 2013; (20):[1-144 pp.]. Available from: http://www.ebrsr.com/evidence-review/20-outcome-measures-stroke-rehabilitation.
7. Indredavik B, Salvesen R, Næss H, Thorsvik D. Nasjonal retningslinje for behandling og rehabilitering ved hjerneslag. In: Helsedirektoratet, editor. 2 ed. Oslo: Helsedirektoratet; 2017. p. 146-62.
8. Johansen I. Outcome of primary health care rehabilitation of older disabled people in two different settings - an open, prospective, comparative observational study. Oslo: University of Oslo; 2013.
9. Dolva A-S, Sveen U, Bonsaksen T, Hagby C, Horghagen S, Solbakken AI, et al. Hvilke undersøkelses- og vurderingsredskaper bruker norske ergoterapeuter? ; resultater fra medlemsundersøkelsen i 2013. Ergoterapeuten. 2015;58(2):44-9.
10. Sunnaas Rehabilitation Hospital. Sunnaas ADL Index [Internet]: Sunnaas Rehabilitation Hospital; 2016 [updated 2016 Oct 19; cited 2017 Sep 09]. Available from: https://www.sunnaas.no/fag-og-forskning/fagstoff/sunnaas-adl-index
11. Vardeberg K, Kolsrud M, Laberg T. Sunnaas index of ADL. J World Federation of Occupational Therapists Bulletin1991. p. 30-5.
12. Sunnaas Rehabilitation Hospital. Scoring sheet - Sunnaas ADL Index www.sunnaas.no2016 [updated 16.07.2021. Available from: https://www.sunnaas.no/Documents/Sunnaas%20ADL-index%20scoreboard%20-%20ENG.pdf.
13. Wales K, Lannin NA, Clemson L, Cameron ID. Measuring functional ability in hospitalized older adults: a validation study. Disabil Rehabil. 2018;40(16):1972-8.
14. Nielsen B, Wæhrens EE. Sunnaas ADL-index - en afprøvning. Ergoterapeuten (danish). 1992;54(10):10-4.
15. Vardeberg K. Prosjektrapport – «Sunnaas ADL-index». Nesoddtangen: Sunnaas Rehabilitation Hospital; 1993.
16. Bathen T, Vardeberg K. Test-Retest Reliability of the Sunnaas ADL Index. Scandinavian Journal of occupational therapy. 2001;8(3):140-7.
17. Kling C, Persson A, Gardulf A. The ADL ability and use of technical aids in persons with late effects of polio. Am J Occup Ther.

2002;56(4):457-61.

18. Wekre LL, Froslie KF, Haugen L, Falch JA. A population-based study of demographical variables and ability to perform activities of daily living in adults with osteogenesis imperfecta. Disabil Rehabil. 2010;32(7):579-87.

19. Korpelainen JT, Niilekselä E, V MV. The Sunnaas Index of activities of daily living: responsiveness and concurrent validity in stroke. Scandinavian Journal of Occupational Therapy. 1997;4(1-4):31-6.

20. Lund A. Inter-rater reliabilitetstest: Barthel og Sunnaas ADL-indeks. Ergoterapeuten (norwegian). 1993;36(7):14-9.

21. Claesson L, Svensson E. Measures of order consistency between paired ordinal data: application to the Functional Independence Measure and Sunnaas index of ADL. Journal of rehabilitation medicine. 2001;33(3):137-44.

22. Olsson B-L, Tervald B. Sunnaas ADL-index: en kritisk granskning. Arbetsterapeuten. 1989;3(12):5-11.

23. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? Journal of Rehabilitation Medicine. 2012;44(2):97-8.

24. Tasmania Uo. Teaching & Learning - Criterion Referenced Assessment: University of Tasmania; 2022 [updated 14.04.2022. Available from: https://www.teaching-learning.utas.edu.au/assessment/criterion-referenced-assessment.

25. Mallinson T. Why Measurement Matters for Measuring Patient Vision Outcomes. Optometry and Vision Science. 2007;84(8):E675-E82.

26. Otterlei G, Stensrud TJ, Feragen A, Horn SA, Schjølberg Å, Ingebretsen GS, et al. Qualilty Report 2017. Sunnaas Rehabilitation Hospital; 2018.

27. Linacre JM. Sample Size and Item Calibration [or Person Measure] Stability. Rasch Measurement Transactions. 1994;7(4):328.

28. REC, Regional Comittees for Medical and Health Research Ethics. Health Research [Internet]: REC - Regional Comittees for Medical and Health Research Ethics; 2012 [cited 2019 Jan 26]. Available from: https://helseforskning.etikkom.no/reglerogrutiner/soknadsplikt?p_dim=34997&_ikbLanguageCode=us.

29. Linacre JM. Optimizing rating scale category effectiveness. Journal of applied measurement. 2002;3(1):85-106.

30. Bond TG, Fox CM. Applying the Rasch model : fundamental measurement in the human sciences. 2nd ed. ed. New York: Routledge; 2012.

31. Linacre JM. Dimensionality: contrasts & variances [Internet]2019 [cited 2019 Jan 28]. Available from: https://www.winsteps.com/winman/principalcomponents.htm.

32. Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the multiple sclerosis impact scale (MSIS-29). Health and Quality of Life Outcomes. 2009;7(1):1-10.

33. Linacre JM. What do Infit and Outfit, Mean-square and Standardized mean? Rasch Measurement Transactions. 2002;16(2):878.

34. Linacre JM. Reliability and separation of measures [Internet]2019 [cited 2019 Jan 24]. Available from: https://www.winsteps.com/winman/reliability.htm.

35. Wright BD, Masters G. Number of Person or Item Strata. Rasch Measurement Transactions. 2002;16(3):1.

36. Linacre JM. DIF Sample Size Nomogram. Rasch Meas Trans. 2013;26(4):1392-402.

37. Norsk Hjerneslagregister. Resultater - 2017 [Internet]: Nasjonalt Servicemiljø for Medisinske Kvalitetsregistre; 2018 [cited 2019 Mar 29]. Available from: https://www.kvalitetsregistre.no/registers/353/resultater/980.

38. Folkehelseinstituttet. Hjerte- og karregisterets statistikkbank [Statistics bank]. [Internet]: Folkehelseinstituttet; 2018 [cited 2019 Jan 22]. Available from: http://statistikkbank.fhi.no/hkr/.

39. Fisher WP. Rating Scale Instrument Quality Criteria www.rasch.org2018 [updated 2.1.2018. Available from: https://www.rasch.org/rmt/rmt211m.htm.

40. Linacre JM. PCA: Data Variance: Explained, Modeled and Empirical. Rasch Measurement Transactions. 2003;17(3):942-3.

41. Pedersen PM, Jørgensen HS, Nakayama H, Raaschou HO, Olsen TS. The Impact of Aphasia on ADL and Social Activities After Stroke: The Copenhagen Stroke Study. Journal of Neurologic Rehabilitation. 1996;10(2):91-6.

42. Sunnaas Rehabilitation Hospital. Sunnaas ADL Index Scoring Manual [Manual]. [Internet]: Sunnaas Rehabilitation Hospital; 2000. Available from: https://www.sunnaas.no/Documents/sunnaas%20adl%20index%20manual%20engelsk.pdf.

43. Byeon H, Koh HW. The relationship between communication activities of daily living and quality of life among the elderly suffering from stroke. Journal of physical therapy science. 2016;28(5):1450-3.

44. Linacre JM, Heinemann AW, Wright BD, Granger CV, Hamilton BB. The structure and stability of the Functional Independence Measure. Archives of physical medicine and rehabilitation. 1994;75(2):127-32.

45. Crişan DR, Tendeiro JN, Meijer RR. Investigating the Practical Consequences of Model Misfit in Unidimensional IRT Models. Applied psychological measurement. 2017;41(6):439-55.

46. Lee H, Lee Y, Choi H, Pyun SB. Community Integration and Quality of Life in Aphasia after Stroke. Yonsei medical journal. 2015;56(6):1694-702.

47. Linacre JM. Largest residual correlations for items [Internet]: Winsteps Help; 2019 [cited 2019 May 07]. Available from: https://www.winsteps.com/winman/table23_99.htm.