Taylor & Francis
Taylor & Francis Group

Disability and Rehabilitation

ISSN: 0963-8288 (Print) 1464-5165 (Online) Journal homepage: www.tandfonline.com/journals/idre20

Derivation and validation of a short form Nottingham extended activities of daily living (SF-NEADL) scale

Alexander Smith, Kalliopi Mavromati, Jonathan Hewitt, Michael Robling & Terence J. Quinn

To cite this article: Alexander Smith, Kalliopi Mavromati, Jonathan Hewitt, Michael Robling & Terence J. Quinn (16 Jan 2026): Derivation and validation of a short form Nottingham extended activities of daily living (SF-NEADL) scale, Disability and Rehabilitation, DOI: 10.1080/09638288.2026.2614225

To link to this article: https://doi.org/10.1080/09638288.2026.2614225

DISABILITY AND REHABILITATION
https://doi.org/10.1080/09638288.2026.2614225

Taylor & Francis
Taylor & Francis Group

ASSESSMENT PROCEDURE

OPEN ACCESS Check for updates

Derivation and validation of a short form Nottingham extended activities of daily living (SF-NEADL) scale

Alexander Smith[a*], Kalliopi Mavromati[b*], Jonathan Hewitt[c], Michael Robling[d] and Terence J. Quinn[b]

[a]Faculty of Medicine, Health and Life Science, Swansea University, Swansea, United Kingdom; [b]School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, United Kingdom; [c]Division of Population Medicine, Cardiff University, Cardiff, United Kingdom; [d]Centre for Trials Research, Cardiff University, Cardiff, United Kingdom

ABSTRACT

Purpose: The Nottingham Extended Activities of Daily Living (NEADL) assessment is commonly used in research and clinical contexts. However, there are concerns surrounding psychometric properties, and with 22-items, NEADL may be too long for clinical use at scale. We aimed to derive a psychometrically robust short form NEADL.

Methods: Data were from the Virtual International Stroke Trials Archive, including individual participant data from 3,6,12 months. Six-month data were used to evaluate NEADL reliability and validity. Corrected item-total correlations identified items for inclusion in the short form (SF-NEADL). The resulting SF-NEADL was then assessed at all time-points for reliability, structural and construct validity, including confirmatory factor analysis (CFA).

Results: NEADL had high internal consistency, and five items with corrected item-total correlations over 0.7 were selected to create a SF-NEADL. The NEADL and SF-NEADL at 6 months had excellent reliability, and construct validity. SF-NEADL reliability and validity were stable at 3 and 12 months. CFA did not suggest unidimensionality of NEADL or SF-NEADL, but SF-NEADL achieved good fit with a two-item structure.

Conclusion: Reliability and validity of our SF-NEADL suggest it is a robust alternative to standard eADL assessments. Its use of fewer and more relevant items makes it suitable for use in busy healthcare settings.

IMPLICATIONS FOR REHABILITATION

- Assessment of ability in extended activities of daily living (eADL) is a fundamental part of research and clinical practice.
- We derived a short form of the Nottingham eADL scale, containing 5 questions about mobility and kitchen tasks, that captures functional independence in daily life as robustly as the original scale.
- With 5 items rather than the original 22, the SF-NEADL is easier to administer and less likely to induce participant fatigue and incomplete response, making it suitable for inclusion in a battery of tests as part of a research or clinical protocol.

ARTICLE HISTORY
Received 2 February 2025
Revised 4 January 2026
Accepted 5 January 2026

KEYWORDS
ADL; activities of daily living; eADL; function; psychometrics; stroke

Introduction

The term "activities of daily living" (ADL) encompasses a wide range of tasks, which are typically distinguished into two groups [1]. Basic activities of daily living (bADLs) are tasks required to survive and meet an individual's basic personal needs. Examples of bADL include ambulating, dressing, and personal hygiene. The Extended or instrumental ADL (eADL or iADL) concept captures those higher-level tasks required to live independently in society, typically including household cleaning, managing finances, and transportation. These tasks come with greater physical and cognitive demands than bADLs [2], thus eADL assessment is used to inform care-needs, to develop rehabilitation goals, and to apply certain diagnoses, for example distinguishing mild cognitive impairment from dementia [3].

CONTACT Terence J Quinn Terry.quinn@glasgow.ac.uk School of Cardiovascular and Metabolic Health, New Lister Building Campus University of Glasgow; Glasgow, UK.
*Joint first authors.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09638288.2026.2614225.

There are many eADL assessment tools available to clinicians and researchers, and no consensus on the optimal tool for a given indication [4]. One of the most commonly used eADL measure in stroke research [5] is the Nottingham Extended Activity of Daily Living scale (NEADL, 6]. NEADL was originally developed for use with stroke survivors but is now applied across many differing healthcare contexts. The NEADL consists of 22 self-report items ordered under 4 subscales, "Mobility", "Kitchen", "Domestic", and "Leisure" with 5 or 6 items per subscale. As well as stroke, NEADL has been validated in various other cohorts, including individuals with multiple sclerosis, and undergoing hip replacement surgery [7–9].

The NEADL is attractive as an assessment for many reasons. NEADL is used for both clinical and research purposes, the content is freely available, and it does not make gender role assumptions seen in other eADL tools. Where psychometric properties have been assessed, the NEADL shows reasonable properties and certainly is comparable to other similar tools for assessing activity limitation [4].

However, the NEADL is not a perfect measure, and there could be scope to further improve its application. While early evaluations of the scale evidenced its validity and test–retest reliability [6,10], more recent evidence has questioned the unidimensionality of the total score [11], and adherence to Guttman scaling principles [9]. Additionally, the tool's high internal consistency [9,12] suggests potential redundancy, and that psychometric performance could be retained with item reduction. While validated short forms of basic ADL scales have been described [23], there is no equivalent short-form EADL scale. Based on familiarity, availability and properties of the original version, the NEADL would seem suited to the development of a short-form assessment that could have clinical and research traction.

The potential utility of NEADL item reduction is further indicated by its application in clinical and research settings. When NEADL is used as a patient-reported outcome measure (PROM), completion rates, although good [13,14], are lower than completion rates when responses are from caregiver proxies [15] suggesting issues with feasibility of scale completion. An issue that may especially seen in those individuals with functional dependence. Thus, as well as reducing general burden on the test subject, shortening the scale could avoid completion biases. Alleviating unnecessary burden on respondents is especially relevant in situations where NEADL may be part of a battery of test measures with associated potential for response fatigue [16,17].

The number of items in a test, or battery of tests, can be reduced, so long as this does not compromise the overall psychometric properties [18]. Indeed, total item reduction has been evidenced to improve rates of completion and reduce missingness [16]. As seen with other scales, reducing the NEADL items while retaining validity has the potential to aid its use.

Our hypothesis was that a psychometrically robust, but shorter assessment of extended ADL could be made using the most discriminating items from the NEADL. To achieve this, we used secondary analysis of data from stroke cohorts, as this is in keeping with the original derivation of the NEADL and represents a population with a range of functional ability. Our specific aims were:

- To examine the properties of the NEADL, and ensure it was psychometrically robust for short-form derivation.
- To derive a short form of the NEADL (SF-NEADL)
- To test the properties of our new short-form assessment.

Methods

Dataset

Data were sourced from the Virtual International Stroke Trials Archive (VISTA, 19). VISTA is a not-for-profit repository for stroke trial data, containing study-quality anonymised individual participant-level data. All studies contained within VISTA, and associated secondary analyses, have been approved by an institutional review committee.

The dataset for this analysis came from UK trials with stroke survivor completed NEADL at short term (3–4 months), medium term (6-months), and longer term (9–12 months) post-stroke. For the derivation of the SF-NEADL, data were from the commonly used research time-point of 6 months post stroke, as this

represents a period of relative stability in stroke functional recovery [20]. Data from other time points were used to assess NEADL and SF-NEADL properties.

Scoring procedure

The original NEADL binary scoring was used, in which the first two responses of 'with help/no' are scored 0 and the second two responses "on my own/on my own with difficulty" are scored 1. Data which were originally scored in alternative formats for example utilizing all four response options per question, were transformed to binary scoring for analysis. Total scores were computed for all participants for total NEADL and at the level of each sub-scale.

Statistical analysis

Assessing NEADL and deriving the short-form. Psychometric descriptives were produced for the original total NEADL, its four subscales, and the derived SF-NEADL, at 6 months. A threshold of over 15% at either tail of the response distribution was used to identify floor and ceiling effects [21].

In deriving the short form, we set a-priori rules of aiming for two items per factor, reducing the number of factors if data allowed, using inter-item and item-total correlations to define the final set of items for inclusion.

Inter-item correlation coefficients were described, where values under 0.2 were considered potentially redundant due to irrelevance to the core construct, and item values over 0.7 were considered potentially redundant due to measuring the same or overlapping constructs [22,23].

Item-total correlation was assessed where values over >0.7 were considered highly correlated [24]. While questions with high item-total correlation can be redundant in a complete scale, this association is a strength during item reduction as it safeguards the internal consistency of the final selection of items [24].

Assessment of reliability and validity. Reliability was measured at scale, and subscale level, for the original NEADL and SF-NEADL at all time-points utilizing Cronbach's alpha coefficient, with values >0.7 considered as an acceptable indication of internal consistency [24].

Validity was evaluated at scale level for NEADL and SF-NEADL at all time points. Construct validity [25] was first analyzed utilizing Spearman's rank correlation coefficient [26] comparing the original total score with the Barthel Index (BI, 27] total score and modified Rankin Scale score (mRS, UKTIA study group, 1988; 28]. The BI as a construct of functional independence in basic activities of daily living and mRS as a measure of global disability were considered suitable for assessing validity, as the two constructs are respectively anticipated to correlate positively and negatively to eADL. Additionally, the SF-NEADL was validated against the concurrent criterion 22-item NEADL at 3–4, 6, and 9–12 months. All Spearman correlation coefficients <0.5 were considered poor, and excellent if above 0.75 [29].

To further evaluate the construct validity *via* assessing the structure captured by the scale, we performed a Confirmatory Factor Analysis (CFA). There is debate in the literature regarding the underlying structure of the NEADL, and a single "eADL" factor has not been definitively proven in previous analyses [11]. So, in our analyses we approached the CFA following the underlying four construct structure for nEADL, then explore the possibility of reduction to two constructs for our short form. Analyses were performed at scale level and then repeated at subscale level for the original NEADL and SF-NEADL at all 3 time points. All analyses were conducted utilizing diagonally weighted least squares (WLSMV), which is recommended for CFA of ordinal data [30]. Absolute model fit was explored in three ways. Firstly, using χ^2, with statistical significance ($p<=0.05$) indicating model fit. Secondly, through Root Mean Square Error of Approximation (RMSEA), where a value of <0.06 is indicative of acceptable fit [31]. Thirdly, *via* Standardized Root Mean Square Residual (SRMR), with a value of <0.08 suggesting an acceptable level of fit. Relative fit was analyzed *via* both the Comparative Fit Index (CFI) and the Tucker–Lewis Index (TLI), with values >0.95 on either measure as indicating good fit [31]. Where both absolute and relative fit is achieved, the former is prioritized, as it is more difficult to achieve as a direct evaluation of the model, not in comparison to a null model [32,33].

As we used a secondary analysis of existing data, we were constrained in the sample size available. Accepting there is no consensus on the approach to sample size calculations in published evaluations

using confirmatory factor analyses, based on the relative simplicity of the model, and lack of missing data, we are confident we had sufficient power for the analysis presented. Certainly our sample size substantially exceeded the 200 participants generally recommended for such analyses [18].

Software. Descriptive and classical test theory analyses of reliability and validity were conducted using SPSS version 27 [34]. All CFA were conducted using MPlus version 8.1 [35].

Results

The six-month post stroke cohort ($N = 722$) consisted of 55% males ($N = 398$), with a median age of 72 years (range:29–92, Table 1). There were no missing or unusable data in the dataset that we accessed.

Floor and ceiling effects were apparent at the subscale level but not for NEADL total score (Floor 4%, Ceiling 6%, Table 2). Floor effects were observed for Mobility (23%), and Domestic (37%) subscales, with ceiling effects apparent in the Mobility (18%), and Kitchen (39%) subscales. The inter-item correlation for

Table 1. Age distribution of the 6-month VISTA cohort.

Age	Frequency	Percent
<60	99	14%
60–69	196	27%
70–79	284	39%
80–89	135	19%
>90	8	1%
Total	722	100%

Table 2. NEADL at 6 months item response distribution and corrected item-total correlation.

		Response counts		
Subscale	Item	0	1	CITC
Mobility (Median = 2, Min = 0, Max = 6, SD ± 2.3)	Do you walk around outside?	337 (47%)	385 (53%)	.698
	Do you climb stairs?	352 (49%)	370 (51%)	.623
	Do you get in and out of the car?	351 (49%)	371 (34%)	.694
	Do you walk over uneven ground?	407 (56%)	315 (44%)	**.715**
	Do you cross roads?	446 (62%)	276 (38%)	**.749**
	Do you travel on public transport?	544 (75%)	178 (25%)	.658
Kitchen (Median = 4, Min = 0, Max = 4, SD ± 1.9)	Do you manage to feed yourself?	142 (20%)	580 (80%)	.459
	Do you manage to make yourself a hot drink?	229 (32%)	493 (68%)	.670
	Do you take hot drinks from one room to the other?	303 (42%)	419 (58%)	**.711**
	Do you do the washing up?	**319 (44%)**	**403 (56%)**	**.701**
	Do you make yourself a hot snack?	368 (51%)	354 (49%)	**.722**
Domestic (Median = 1, Min = 0, Max = 5, SD ± 1.7)	Do you manage your own money when you are out?	317 (44%)	405 (56%)	.678
	Do you wash small items of clothing?	456 (63%)	266 (37%)	.639
	Do you do your own housework?	533 (74%)	189 (26%)	.648
	Do you do your own shopping?	564 (78%)	158 (22%)	.665
	Do you do a full clothes wash?	521 (72%)	201 (28%)	.641
Leisure (Median = 2, Min = 0, Max = 6, SD ± 1.6)	Do you read newspapers or books?	151 (21%)	571 (79%)	.397
	Do you use the telephone?	178 (25%)	544 (75%)	.511
	Do you write letters?	453 (63%)	269 (37%)	.504
	Do you go out socially?	528 (73%)	194 (27%)	.664
	Do you manage your own garden?	605 (84%)	117 (16%)	.528
	Do you drive a car?	629 (86%)	93 (13%)	.441
Total (Median = 9, Min = 0, Max = 22, SD ± 6.7)				

Note. Responses "no" or "with help" were scored 0, while responses "on my own with difficult" or "independently" were scored 1. Corrected item-total correlations (CITC) in bold were over .7 and included in the SF-NEADL.

all 22 items highlighted one pair of items correlated over the >0.7 threshold for potential redundancy: "Do you manage to make yourself a hot drink?" and "Do you take hot drinks from one room to the other?".

The following five items exceeded the predetermined threshold of 0.7 corrected item-total correlation, but had inter-item correlations that did not suggest redundancy and were selected for inclusion in the SF-NEADL: "Do you walk over uneven ground?"; "Do you cross roads?"; "Do you take hot drinks from one room to the other?"; "Do you do the washing up?"; "Do you make yourself a hot snack?". (Table 2, Supplemental Materials A) The first two items are from the original tool's Mobility subscale, while the last three are from the Kitchen subscale. While the "Domestic" subscale included items with suitable corrected item-total correlations, inter-item correlations for the "Domestic" and "Leisure" subscale were not favorable, and in keeping with our development rules we opted to restrict the short form to the remaining two subscales. Response distributions (Table 3) evidenced floor and ceiling effects over 15% at subscale level and total score.

The derived SF-NEADL was also calculated using data from the 3–4 months ($N = 305$) and 9–12 months ($N = 535$) responses to the complete scale. Reliability as measured *via* internal consistency (Table 4) was highest overall and across subscales in the original NEADL. Across time points, SF-NEADL reliability was acceptable at subscale level and close to excellent overall.

Convergent validity (Table 5) with Barthel Index was excellent for the NEADL at 6 months and the SF-NEADL at all time points, while for the mRS validity was acceptable but not excellent for the NEADL at 6 months and for the SF-NEADL at all available time points. Correlation of SF-NEADL was strong with the concurrent NEADL criterion at 3–4, 6, and 9–12 months.

Table 3. SF-NEADL at 6 months response by subscale in original scale.

		Response Counts	
Subscale	Item	0	1
Mobility (Median = .00, Min = 0, Max = 2, SD ± .9)	Do you walk over uneven ground?	407 (56%)	315 (44%)
	Do you cross roads?	446 (62%)	276 (38%)
Kitchen (Median =2, Min = 0, Max = 3, SD ± 1.3)	Do you take hot drinks from one room to the other?	303 (42%)	419 (58%)
	Do you do the washing up?	319 (44%)	403 (56%)
	Do you make yourself a hot snack?	368 (51%)	354 (49%)
Total (Median = 2, Min = 0, Max = 5, SD ± 2)			

Note. Responses "no" or "with help" were scored 0, while responses "on my own with difficult" or "independently" were scored 1.

Table 4. Internal consistency reliability for NEADL and SF-NEADL.

		Cronbach's Alpha				
Scale	Months post stroke	Overall	Mobility	Kitchen	Domestic	Leisure
NEADL	6	.940	.889	.866	.836	.734
SF-NEADL	3–4	.852	.753	.874	–	–
	6	.863	.781	.811		
	9–12	.860	.791	.849	–	–

Table 5. Correlation coefficients for validity assessments.

		Spearman Correlation Coefficient		
Scale	Months post-stroke	Barthel	mRS	Total NEADL at 6 months
NEADL	6	r_s (699) = .806, $p < .0005$	r_s (722) = −0.714, $p < .0005$	–
SF-NEADL	3–4	r_s (282) = .790, $p < .0005$	mRS data unavailable	r_s (305) = .927, $p < .0005$
	6	r_s (699) = .787, $p < .0005$	r_s (349) = −0.656, $p < .0005$	r_s (722) = .944, $p < .0005$
	9–12	r_s (373) = .794, $p < .0005$	r_s (301) = −0.699, $p < .0005$	r_s (535) = .941, $p < .0005$

Note. Correlations with Barthel Index (Barthel) and modified Rankin Scale (mRS) are based on data from the corresponding time point. The mRS score (i.e., level of disability) is anticipated to be negatively correlated with NEADL (i.e., functional independence) as they represent inversely related theoretical constructs.

In assessing the structural validity (Supplemental Materials B) of the NEADL, absolute model fit was not possible, when attempting to fit the 4-factor model reflecting the subscales or a unidimensional single "extended ADL". Similarly, the unidimensional single-factor "extended ADL" model failed to achieve absolute model fit for the SF-NEADL at 6 months, however, the 2-factor model (Mobility 2 items, Kitchen 3 items) achieved good fit on all measures of model fit. Standardized item factor loading scores for the 2-factor model all loaded at >0.9.

Discussion

In alignment with previous evidence [11], our findings suggest suboptimal psychometric performance of the original NEADL. Using a validated approach, we derived a five-item version of NEADL designed to ease test burden and facilitate greater completion. Across various measures of reliability and validity, assessed at different time points, our SF-NEADL appeared psychometrically robust.

Our analysis of the original NEADL supported an item reduction approach and by implication, the creation of a shorter form. As previously described [9,12], the NEADL had high internal consistency in keeping with a degree of redundancy within the scale. Items with highest correlations were predominantly contained within the Domestic and Leisure subscales suggesting the importance of these tasks when considering eADL, but also the potential functional equivalence of some of these actions.

While a psychometrically valid approach, our motivation for item reduction was primarily to improve the efficacy of measurement by reducing the amount of time and cognitive effort required of participants. Item reduction should not sacrifice psychometric strength, and our analyses of the SF-NEADL was reassuring in this regard, with consistently high internal consistency, and agreement with related measures demonstrated at various time points. This consistency indicates high external validity of the scale, as it is robust to time-related physical and psychosocial changes in life post stroke. However, reliability and construct validity were highest from six months post stroke onwards, likely due to increasing stability of functional ability over this time [20] and representing epochs where ADL measurement is most common in research. At these time points, our analyses of the SF-NEADL revealed quantitively similar properties to the original scale.

Deriving our SF-NEADL using analysis of NEADL data, revealed important issues associated with the original scale and subsequently its shorter form. Namely, the poor structural validity of the original NEADL raises the question of whether this eADL measure, and its four original subscales, are fit for purpose. The CFA of the original NEADL evidenced relative fit, but not absolute fit, which is a stricter evaluation [32,33]. This is likely a result of the use of Guttman scaling analyses which by nature do not account for the structure of the instrument in establishing interval-level data from ordinal data, but instead assume unidimensionality [36]. The SF-NEADL, however, had good structural validity when absolute and relative fit was assessed in terms of the Mobility and Kitchen subscales.

The observed floor and ceiling effects in the 6-month NEADL further reinforce concerns about the appropriateness of the original scale structure. In eADL measurement, the split of responses is a reflection of the difficulty of each item's task. Therefore, our data suggest that some eADL tasks are experienced as excessively difficult or easy. The ceiling effects are a particular concern for use of the scale in research, as important between-group differences, or improvements may not be captured beyond the scale maximum score. The floor effects are perhaps less of a concern, as should a subject struggle with eADL tasks, the assessor can then test against less demanding basic ADLs. Given floor and ceiling effects in the original scale, it is no surprise that such patterns are still present in our short form, but the improved quantitative validity of the SF-NEADL suggests that the reduction of the breadth of task difficulty facilitates application of the scale.

The item reduction also addressed overlapping items (e.g., making a hot drink and taking it to another room), contemporaneously irrelevant tasks (e.g., writing letters, doing a small load of laundry) or tasks that may have cultural associations with gender roles (e.g., doing one's own housework), which impact the face validity of the scale. Such differences in complexity and difficulty do not go unnoticed by people who are assessed using the scale [37].

Indeed, the improved face and structural validity of our SF-NEADL post item reduction perhaps speaks to a broader issue around the content of eADL measurement (for a review see 4]. The fact that exclusion of the Domestic and Leisure tasks subthemes did not substantially impact the reliability and validity of

the SF-NEADL, in conjunction with our finding that SF-NEADL shaped by the two factors of Mobility and Kitchen had improved structural validity over its unidimensional equivalent, suggests there is a fundamental problem with the operational definition of the construct being measured. In the context of additional evidence that the eADL construct is not unidimensional [11], it remains unclear what exactly we are assessing under the auspices of 'extended ADL'.

There are strengths and limitations to the analyses we present. Both the NEADL evaluation and SF-NEADL derivation were performed with cohorts larger than traditionally included in psychometric evaluations. The secondary nature of the data analyzed suggests that the responses included are a direct reflection of the studies for which the NEADL, and any shorter versions, are intended to be used. Thus, making these results externally valid and hopefully generalizable to real-world use of the SF-NEADL. This analysis followed a standard approach of using corrected-item total correlation for item reduction [24]. Using a single measure for decision-making in this process is not optimal as it has the potential to lower construct validity, yet performance of the scale post-reduction remained acceptable. An important caveat to our analyses, is that the validation of the SF-NEADL was based on data from the complete 22-item original NEADL. Future assessments of SF-NEADL properties should be based on direct use of the 5-item scale.

This limitation suggests an important direction for future research. At present all of our validation has been based on secondary analyses of trial data. We assume that the short form will offer advantages in terms of speed, item completion, and test burden, but we are unable to empirically test this with the data available. Prospective assessments using our SF-NEADL and including measures of test experience and feasibility are needed before the short form can be definitively recommended.

Even with excellent psychometric performance, the SF-NEADL cannot address the fundamental issues with the operational definition and construct structure of functional independence as measured by NEADL. This is likely to be true for all measures of the EADL construct. Future research also needs to be cognizant of the culturally fluid nature of "day to day" tasks. Task inclusion and wording needs to reflect the variety of ways in which an outcome can be achieved, with consideration for accommodations accessible to individuals in the twenty-first century. Underlying socially constructed assumptions about what an independent adult's daily life looks like, what resources are available to them, and the moderating effects of social role, gender and other factors will inevitably shape the tasks included in an eADL measure and the wording of their description. It seems likely that future iterations of eADL measures will use technological approaches that allow for assessments that are more personalized to the individual. However, any such development should be tested with end users, and comparisons made with traditional approaches to ensure that technological sophistication also brings improvement in performance, especially for older adults [38].

Redefining eADLs, and reforming their measurement, is a potentially resource- and time-intensive undertaking, and one for which clinical practice and research cannot wait. As 'big data' become increasingly important, approaches that offer improved efficiency and limit human error, without compromising psychometric properties, are especially relevant. We believe our SF-NEADL speaks to this need. As an example, the ResQ international registry of stroke care is aiming to collect eADL data and hopes to use our SF-NEADL as part of a self-report digital assessment battery [39] Including SF-NEADL in self-report digital assessment batteries may offer opportunities to monitor eADL changes over time and across diverse populations for a fraction of the cost and time required for traditional in-person assessment. However, again, any such intervention should be empirically tested before implementation at scale.

There have been many examples of new assessment scales that, while psychometrically superior to previous iterations, have failed to gain traction. However, short versions of existing scales have proven popular, for example, the 5-min Montreal Cognitive Assessment (MoCA) protocol [27] is now incorporated into guideline-recommended assessment batteries [40]. The increasing use of short-form MoCA has perhaps been aided by ongoing research to describe the properties of the test. We would encourage teams to subject our SF-NEADL to similar examinations in real-world settings.

Ultimately, measuring performance across eADL tasks is likely to remain an important part of clinical and research assessment, capturing functional independence with ecological validity, while preserving an indirect measure of quality of life and psychological well-being [41]. Despite its popularity, the NEADL is limited by item redundancy, poor structural and face validity. As a shorter but equally robust measure our novel 5-item SF-NEADL offers time, test burden and opportunity cost efficiencies.

Acknowledgments

With thanks to all the academic teams and stroke survivors who contributed data to the studies included in the VISTA IPD repository.

The copyright for the Nottingham Extended Activities of Daily Scale is held by University of Nottingham. It is free for clinical and research use, providing the source is cited. The authors of this paper have corresponded with the Nottingham team, and they have approved this exploration of a short form. Further details on use of the scale can be found at https://www.nottingham.ac.uk/medicine/research/research-areas/rehabilitation-and-ageing/published-assessments.aspx.

Disclosure statement

This work was part supported by a Stroke Association PhD Fellowship (grant number SA PGF 18\100029) and by the European Union as a part of the Horizon Europe research initiative RES-Q+ (grant number 101057603). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

ORCID

Alexander Smith http://orcid.org/0000-0001-9656-6751
Kalliopi Mavromati http://orcid.org/0000-0002-6600-064X
Jonathan Hewitt http://orcid.org/0000-0002-7924-1792
Terence J. Quinn http://orcid.org/0000-0003-1401-0181

Data availability statement

The original data are available from the VISTA resource and can be accessed through *via* the online Virtual Trials Archive https://www.virtualtrialsarchives.org/vista/

References

[1] Wu C, Chuang L, Lin K, et al. Responsiveness and validity of two outcome measures of instrumental activities of daily living in stroke survivors receiving rehabilitative therapies. Clin Rehabil. 2011;25(2):175–183. doi:10.1177/0269215510385482.
[2] Sikkes SAM, de Lange-de Klerk ESM, Pijnenburg YAL, et al. A systematic review of instrumental activities of daily living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. 2009;80(1):7–12. doi:10.1136/jnnp.2008.155838.
[3] Wales K, Clemson L, Lannin N, et al. Functional assessments used by occupational therapists with older adults at risk of activity and participation limitations: a systematic review. PLoS One. 2016;11(2):e0147980-e0147980. doi:10.1371/journal.pone.0147980.
[4] Kelbling E, Ferreira Prescott D, Shearer M, et al. An assessment of the content and properties of extended and instrumental activities of daily living scales: a systematic review. Disabil Rehabil. 2024;46(10):1990–1999. doi:10.1080/09638288.2023.2224082.
[5] Quinn TJ, Dawson J, Walters MR, et al. Functional outcome measures in contemporary stroke trials. Int J Stroke. 2009;4(3):200–205. doi:10.1111/j.1747-4949.2009.00271.x.
[6] Nouri F, Lincoln N. An extended activities of daily living scale for stroke patients. Clin Rehabil. 1987;1(4):301–305. doi:10.1177/026921558700100409.
[7] Ferrucci L, Koh C, Bandinelli S, et al. 2010. Disability, functional status, and activities of daily living. In Encyclopedia of gerontology. (pp. 427–436). Elsevier Inc. Netherlands.
[8] Harwood RH, Ebrahim S. The validity, reliability and responsiveness of the Nottingham extended activities of daily living scale in patients undergoing total hip replacement. Disabil Rehabil. 2002;24(7):371–377. doi:10.1080/09638280110101541.
[9] Nicholl CR, Lincoln NB, Playford ED. The reliability and validity of the Nottingham extended activities of daily living scale in patients with multiple sclerosis. Mult Scler. 2002;8(5):372–376. doi:10.1191/1352458502ms827oa.
[10] Lincoln NB, Gladman JR. The extended activities of daily living scale: a further validation. Disabil Rehabil. 1992;14(1):41–43. doi:10.3109/09638289209166426.
[11] das Nair R, Moreton BJ, Lincoln NB. Rasch analysis of the Nottingham extended activities of daily living scale. J Rehabil Med. 2011;43(10):944–950. doi:10.2340/16501977-0858.

[12] Gladman JRF, Lincoln NB, Adams SA. Use of the extended ADL scale with stroke patients. Age Ageing. 1993;22(6):419–424. doi:10.1093/ageing/22.6.419.
[13] Murray J, Forster A, Young J. Response and completion rates for postal outcomes booklets in stroke rehabilitation. Int J Ther Rehabil. 2007;14(10):440–445. doi:10.12968/ijtr.2007.14.10.27395.
[14] Parker CJ, Gladman JR, Drummond AE, et al. A multicentre randomized controlled trial of leisure therapy and conventional occupational therapy after stroke. Clin Rehabil. 2001;15(1):42–52. doi:10.1191/026921501666968247.
[15] Logan P, Ahern J, Gladman J, et al. A randomized controlled trial of enhanced social service occupational therapy for stroke patients. Clin Rehabil. 1997;11(2):107–113. doi:10.1177/026921559701100203.
[16] Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. Value Health. 2011;14(8):1101–1108. doi:10.1016/j.jval.2011.06.003.
[17] Turner RR, Quittner AL, Parasuraman BM, et al. Patient-reported outcomes: instrument development and selection issues. Value Health. 2007;10 Suppl 2: s 86–S93. doi:10.1111/j.1524-4733.2007.00271.x.
[18] Kyriazos T. Applied Psychometrics: sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. PSYCH. 2018;09(08):2207–2230. doi:10.4236/psych.2018.98126.
[19] Ali M, Bath PMW, Hennerici MG, et al. The virtual international stroke trials archive. Stroke. 2007;38(6):1905–1910. doi:10.1161/STROKEAHA.106.473579.
[20] Lee KB, Lim SH, Kim KH, et al. Six-month functional recovery of stroke patients: a multi-time-point study. Int J Rehabil Res. 2015;38(2):173–180. doi:10.1097/MRR.0000000000000108.
[21] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293–307. doi:10.1007/bf01593882.
[22] Clark LA, Watson D. Constructing validity: basic issues in objective scale development. Psychol Assess. 1995;7(3):309–319. doi:10.1037/1040-3590.7.3.309.
[23] Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–55. doi:10.5116/ijme.4dfb.8dfd.
[24] Boateng GO, Neilands TB, Frongillo EA, et al. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health. 2018;6:149–149. doi:10.3389/fpubh.2018.00149.
[25] Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. J Adv Nurs. 2014;70(9):1970–1979. doi:10.1111/jan.12402.
[26] Buntragulpoontawee M, Phutrit S, Tongprasert S, et al. Construct validity, test–retest reliability, and internal consistency of the Thai version of the Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH-TH) in patients with carpal tunnel syndrome. BMC Res Notes. 2018;11(1):208. doi:10.1186/s13104-018-3318-5.
[27] Mahoney FI, Barthel DW. functional evaluation: the Barthel index. Md State Med J. 1965;14:61–65.
[28] Rankin J. Cerebral vascular accidents in patients over the age of 60. II. prognosis. Scott Med J. 1957;2(5):200–215. doi:10.1177/003693305700200504.
[29] Portney LG, Watkins MP. ; 2009. Foundations of clinical research: applications to practice. (Vol. 892, pp. 11–15. Upper Saddle River, NJ: pearson/Prentice Hall.
[30] Li C. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods. 2016;48(3):936–949. doi:10.3758/s13428-015-0619-7.
[31] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling. 1999;6(1):1–55. doi:10.1080/10705519909540118.
[32] Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research Online. 2003;8(2):23–74.
[33] Tabachenik D, Fidel J. Structural equation modeling: guidelines for determining model fit. J Bus Res Methods. 2012;6:1–55.
[34] IBM Corp. IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp. New York, USA 2020.
[35] Muthén LK, Muthén BO. Mplus User's Guide. (8th ed.). Muthén & Muthén; Los Angeles, USA. 1998–2017.
[36] Andrews JA, Peterson M. The development of social images of substance users in children: a guttman unidimensional scaling approach. J Subst Use. 2006;11(5):305–321. doi:10.1080/14659890500419774.
[37] Minnis M, Burton JK, Kelbling E, et al. Not daily, sometimes not ever – mixed methods exploration of the contemporary relevance of tasks contained in extended activities of daily living scales. Age Ageing. 2024;53(8). doi:10.1093/ageing/afae185.
[38] Goodwin VA, Low SAM, Quinn TJ, et al. Including older people in health and social care research: best practice recommendations based on the INCLUDE framework. Age Ageing. 2023;52(6):afad082. doi:10.1093/ageing/afad082.
[39] Fasugba O, Sedani R, Mikulik R, et al. How registry data are used to inform activities for stroke care quality improvement across 55 countries: a cross-sectional survey of registry of stroke care quality (RES-Q) hospitals. Eur J Neurol. 2024;31(1):e16024-n/a. doi:10.1111/ene.16024.
[40] Hachinski V, Iadecola C, Petersen RC, et al. National institute of neurological disorders and stroke-canadian stroke network vascular cognitive impairment harmonization standards. Stroke (1970). 2006;37(9):2220–2241. doi:10.1161/01.STR.0000237236.88823.47.
[41] Beltrami LPB, Marques PT, Barbosa FJL, et al. Functional impairment and post-stroke depression: a 6-month longitudinal study. Trends Psychiatry Psychother. 2025;47:e20220589. doi:10.47626/2237-6089-2022-0589.