

Research Paper

Evaluation of an activities of daily living scale for adolescents and adults with developmental disabilities

Matthew J. Maenner, Ph.D.^{*}, Leann E. Smith, Ph.D., Jinkuk Hong, Ph.D., Renee Makuch, B.S., Jan S. Greenberg, Ph.D., and Marsha R. Mailick, Ph.D.

Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Room 529A, Madison, WI 53705, USA

Abstract

Background: Activity limitations are an important and useful dimension of disability, but there are few validated measures of activity limitations for adolescents and adults with developmental disabilities.

Objective/hypothesis: To describe the development of the Waisman Activities of Daily Living (W-ADL) Scale for adolescents and adults with developmental disabilities, and systematically evaluate its measurement properties according to an established set of criteria.

Methods: The W-ADL was administered among four longitudinally studied groups of adolescents and adults with developmental disabilities: 406 with autism; 147 with fragile-X syndrome; 169 with Down syndrome; and 292 with intellectual disability of other or unknown origin. The W-ADL contains 17 activities and each is rated on a 3-point scale (0 = “does not do at all”, 1 = “does with help”, 2 = “independent”), and a standard set of criteria were used to evaluate its measurement properties.

Results: Across the disability groups, Cronbach’s alphas ranged from 0.88 to 0.94, and a single-factor structure was most parsimonious. The W-ADL was reliable over time, with weighted kappas between 0.92 and 0.93. Criterion and construct validity were supported through substantial associations with the Vineland Screener, need for respite services, caregiving burden, and competitive employment. No floor or ceiling effects were present. There were significant group differences in W-ADL scores by maternally reported level of intellectual disability (mild, moderate, severe, profound).

Conclusions: The W-ADL exceeded the recommended threshold for each quality criterion the authors evaluated. This freely available tool is an efficient measure of activities of daily living for surveys and epidemiological research concerning adolescents and adults with developmental disabilities. © 2013 Elsevier Inc. All rights reserved.

Keywords: Activities of daily living; Adolescent; Adult; Autistic disorder; Developmental disabilities; Down syndrome; Fragile X syndrome; Intellectual disability

Abbreviations: ADL, activities of daily living; ICC, intra-class correlation coefficient; ICF, International Classification of Functioning, Disability, and Health; SEM, standard error of the measure; SDC, smallest detectable change; W-ADL, Waisman Activities of Daily Living (Scale).

This study was presented in abstract/poster form at the 45th Annual Gatlinburg Conference on March 7, 2012, and at the International Meeting for Autism Research on May 18, 2012.

This manuscript was supported by grants from the National Institute on Aging (R01 AG08768, M.R. Mailick, PI) and from the National Institute of Child Health and Human Development to the IDDRC at the University of North Carolina (P30 HD003100-S1) to support a Fragile X Research Center. The Fragile X Research Center has three additional sites (Research Triangle Institute International, the University of Wisconsin-Madison, and University of Kansas). The present analysis was based on data collected at the UW-Madison Waisman Center (T32 HD07489, M.R. Mailick, PI) site. We also acknowledge the support we received from the Waisman Center (P30 HD03352, M.R. Mailick, PI).

Conflicts of interest: None.

* Corresponding author. Tel.: +1 608 890 0752.

E-mail address: mjmaenner@wisc.edu (M.J. Maenner).

Developmental disabilities are heterogeneous and life-long disorders that are often characterized by problems related to the functioning of the brain or senses and include genetic disorders that affect cognition, behavior, and multiple body systems.^{1–3} The causes and consequences of these conditions may vary widely, even within a particular ‘type’ of disability. For example, some adults with autism experience severe limitations in basic tasks necessary for independent living, while others have no limitations in these areas (but may still struggle in social situations).⁴ The nature and severity of these limitations are important for providing appropriate services and interventions, for research, and for informing public health policy.^{5,6} There are, however, few freely available and high-quality tools for measuring activities among adolescents and adults with substantial developmental disabilities. We sought to evaluate a new daily activities measure for this population that is feasible for surveys and epidemiological research.

Several widely used theoretical frameworks of disability account for and describe aspects of disability beyond a medical classification.^{7–9} The World Health Organization's International Classification of Functioning, Disability, and Health (ICF) describes disability at three different levels: impairments (problems or deviations in the structure or functioning of a body part), activity limitations (difficulty executing tasks or actions), and participation restrictions (problems with involvement in life situations).¹⁰ An innovative aspect of the ICF model is the recognition that both personal and environmental factors may affect the experience of disability. Importantly, the ICF distinguishes between “performance” (what an individual does in his or her current environment) and “capacity” (an individual's ability to perform a task or action, the highest probable level of functioning) for activities and participation.¹⁰

A linear or consequential relationship between different levels of disability (e.g., from impairment to activity limitations) is neither presumed nor inevitable; however, adolescents and adults with developmental disabilities are likely to experience activity limitations.¹¹ In addition to the variability within diagnostic categories for many developmental disabilities, other co-occurring health conditions, social support, and the opportunities and constraints of the physical environment may contribute to an individual's performance of activities. Measurement of the complex relationships between different aspects of disability is greatly aided through the use of precise language and a strong conceptual framework, such as that offered by the ICF.

Particularly for adolescents and adults with known developmental disabilities, a greater emphasis on activities offers an opportunity to acknowledge that “what people do is more important than the expectation that they do things ‘normally,’”¹² although this has not been the focus of most available measures. Other problems with available measures include “floor” effects and developmental-level anchoring which limit the usefulness of such measures. Specifically, some individuals with autism or intellectual disability are likely to score at the extreme minimum of population-normed measurement tools, and previous studies have suggested that there are important differences among individuals scoring at the “floor” on standardized measures.^{13–15} Some measures require raters to first determine an individual's “mental” or “developmental” age, and then administer items intended for typically developing individuals of that developmental level. This can result in assessment of adults with developmental disabilities with items that were designed for young children. For adults with substantial limitations, sensitive and age-appropriate measures are needed to detect meaningful changes over time and to distinguish between people with different needs.

Despite the widespread recognition that activity limitations represent an important dimension of disability, we

are not aware of any freely available and thoroughly evaluated tools for measuring activity limitations in adults with developmental disabilities. Activities of Daily Living (ADL) scales have been used extensively in research and clinical applications, often for individuals recovering from surgery or a stroke, or to measure functional decline associated with aging and dementia. Some US states determine service eligibility and financial assistance to adults with limitations on the basis of ADL scores; in 2009, the state of Montana provided disability-related financial support at a rate of \$34 per month for each point scored on a scale of ADL items.¹⁶ ADLs are also considered durable indicators of activity limitations and consistent with the World Health Organization's dimensional framework for disability.¹⁷ Even though ADL-based measures are well-known and presumed to be useful, there is a paucity of rigorous research supporting the reliability and validity of ADL scales.¹⁸ Establishing the quality of an ADL measure intended for populations with developmental disabilities could help articulate the needs of adults with developmental disabilities and facilitate better research and service provision.

The purpose of this paper is to describe the development and evaluate the measurement properties of the Waisman Activities of Daily Living (W-ADL) Scale for adolescents and adults with developmental disabilities. To guide our evaluation, we applied a widely cited set of quality criteria for health measurement scales.¹⁹

Materials and methods

Subjects and studies

This analysis makes use of data collected by three longitudinal studies comprised of four different disability groups (autism, Down syndrome, fragile X, and intellectual disability of unknown etiology), that together include 1014 study participants. Although these studies were not prospectively designed for the development of the W-ADL, they contain a wealth of information related to the health and functioning of adults with developmental disabilities. This analysis utilized relevant items and measures when they were available in each of the studies. The study populations and selected measures used in the present analysis are briefly described here; detailed descriptions have been previously published and are cited below.

Autism

Data regarding 406 adolescents and adults with autism (referring to all autism spectrum disorders) initially were included in an ongoing longitudinal study including 9 points of measurement spanning a 12-year period.²⁰ The median age of the participants at Time 1 was 18 years (range: 10–52), and 73% of the participants were male. By parent report, 6% also had cerebral palsy and 23% had epilepsy or a seizure disorder. Independent autism diagnoses were

confirmed with a research-administered Autism Diagnostic Interview-Revised.²¹ The W-ADL was completed at Times 1, 4, 7, and 8; approximately ten years elapsed between data collection between Times 1 and 8. Data collection for Times 7 and 8 occurred approximately 18 months apart. The Vineland Screener, a measure of adaptive behavior, was administered at Time 4. The Vineland Screener is highly correlated with the full Vineland Scales (correlation coefficients ranging from 0.87 to 0.98).²² Employment and education were coded according to Taylor and Seltzer's Vocational Index for individuals with autism.²³

Fragile X

Data regarding 147 adolescents and adults with fragile X syndrome were included in the first round of data collection in an ongoing longitudinal study.²⁴ The median age at the time of interview was 18 years (range: 12–48), and 82% of the participants were male. By parent report, 2% also had cerebral palsy and 11% had epilepsy or a seizure disorder. Fragile X was ascertained through genetic test results from the participants' medical records. Data from the first time point are currently available. Mothers' perceived caregiving burden for their adolescent or adult child with fragile X was assessed using the Zarit Burden Scale.²⁵

Down syndrome and intellectual disability

Data regarding 169 adolescents and adults with Down syndrome and 292 adolescents and adults with other intellectual disabilities were included in a 10-year longitudinal study.^{26,27} The median age of the participants with Down syndrome at Time 1 was 31 years (range: 15–55), and 61% were male. Down syndrome was ascertained via parent report. The median age of the participants with other intellectual disabilities at Time 1 was 35 years (range: 17–66), and 50% were male. Of the 292 individuals with intellectual disability (but not Down syndrome), maternal report indicated that 23% had cerebral palsy and 13% had epilepsy (4% were reported to have both). Of the 169 with Down syndrome, 1% also had cerebral palsy and 2% had epilepsy or a seizure disorder. Data were collected at eight time points (approximately 18 months apart and more than ten years total between Times 1 and 8), and the W-ADL was given at each time. Mothers were the primary respondents for this study; however, a subset of fathers concurrently responded to W-ADL items at Time 6. Stanford-Binet intelligence tests were administered at Times 2 and 3 by professional clinicians and graduate students in educational psychology.

Evaluation of quality criteria (as adapted from Terwee and colleagues)

Terwee and colleagues¹⁹ proposed eight criteria to comprehensively evaluate the measurement properties of an instrument: content validity, internal consistency,

criterion validity, construct validity, reproducibility, responsiveness, floor or ceiling effects, and interpretability. We summarized the purpose of each of the criteria and applied each criterion to the W-ADL scale using the methodology recommended by Terwee and colleagues.

Results

Content validity

Content validity is established via descriptions of the aim of the measure, target population, item selection and reduction, and an assessment of the measure's interpretability and feasibility. Below, we describe the purpose, development, and feasibility of the W-ADL.

The W-ADL aims to measure the level of independence in performing typical daily activities for adolescents and adults. These activities include dressing, grooming, housework/chores, meal-related activities, and activities outside the home. The target population is adolescents and adults with substantial intellectual and/or developmental disabilities (including autism, fragile X, Down syndrome, and intellectual disability of unknown or other etiologies). The W-ADL is intended to show changes in daily activities over time, or differences between groups among people with substantial disabilities. Possible applications include measuring the natural history and trajectories of daily activities for people with a particular disability, measuring outcomes of a program designed to improve independence in daily living skills, or helping to identify environmental or social factors that promote the performance of ADLs.

Members of the research team, which included experienced clinicians and family members of individuals with developmental disabilities, selected and modified items from earlier ADL questionnaires (such as the Barthel²⁸ and Lawton²⁹ scales) and identified additional relevant activities. The W-ADL is a criterion-referenced measure, as it is designed to provide information about the level of performance of common daily activities for adults. Researchers adapted items to be succinct, appropriately worded, and applicable to the target population. Prior to each wave of data collection, items were pilot tested and feedback was solicited from respondents on the length and appropriateness of the items. Over time, items that were nearly always concordant were collapsed into a single item (such as bathing upper and lower body). We excluded items that were frequently unanswered or items which were especially dependent upon an individual's environment. For example, we excluded questions about public transportation or taxi use because participants frequently indicated that these items were not applicable to their situation.

The 17 items in the final version of the W-ADL (Table 1) pertain to the current or expected performance of the target individual at the time when the survey was given. The target adult's performance of each activity is rated on a 3-point scale (0 = "does not do at all", 1 = "does

Table 1
W-ADL items and scoring

PLEASE RATE (name of child)'S LEVEL OF INDEPENDENCE IN ...	Ability to perform task: 2 = independent or does on own; 1 = does with help; 0 = does not do at all
1. Making his/her own bed	
2. Doing household tasks, including picking up around the house, putting things away, light housecleaning, etc.	
3. Doing errands, including shopping in stores	
4. Doing home repairs, including simple repairs around the house, non-technical in nature; for example, changing light bulbs or repairing a loose screw	
5. Doing laundry, washing and drying	
6. Washing/bathing	
7. Grooming, brushing teeth, combing and/or brushing hair	
8. Dressing and undressing	
9. Toileting	
10. Preparing simple foods requiring no mixing or cooking, including sandwiches, cold cereal, etc.	
11. Mixing and cooking simple foods, fry eggs, make pancakes, heat food in microwave, etc.	
12. Preparing complete meal	
13. Setting and clearing table	
14. Drinking from a cup	
15. Eating from a plate	
16. Washing dishes (including using a dishwasher)	
17. Banking and managing daily finances, including keeping track of cash, checking account, paying bills, etc. (Note: if he/she can do a portion but not all circle '1' with help.)	

Instructions: "We would like to know about your son or daughter's current level of independence in performing activities of daily living. For each activity, please tell me the number which best describes your son/daughter's ability to do the task. For example, independent would mean your son/daughter is able to do the task without any help or assistance."

with help", 2 = "independent"), and item scores are summed to produce an overall score.

Nearly all of the respondents were biological parents of the target adult. W-ADL items were administered at the beginning of each study, at three additional time points for adults with autism, and seven additional time points for adults with Down syndrome or intellectual disability. In the autism group, at four time points, a total of 1087 of 1121 (97.0%) attempted W-ADLs had complete responses to all items. In the Down syndrome and intellectual disability groups, across 8 time points, 2798 of 3023 (92.6%) W-ADLs were completed. At the single time point in the fragile X group, 100% of W-ADLs for 147 participants were completed. On average, we estimate the W-ADL took approximately 5 min for parents to complete.

Internal consistency

Internal consistency is a measure of the relatedness between items, with greater correlation suggesting that items are measuring the same underlying construct. We did not hypothesize potential subscales; rather, we evaluated the item structure using exploratory factor analysis. We also used the R-implementation of Revelle's ICLUST (hierarchical clustering) algorithm to further describe the item structure.^{30,31} We then calculated Cronbach's alpha and evaluated the alpha according to Terwee and colleagues' recommended alpha values of 0.70–0.95.

We performed exploratory factor analysis of the W-ADL items at the first time point for all participants. We observed

a single factor accounting for 42% of the variance, with item loadings ranging from 0.46 to 0.77. Additional factors resulted in marginal increases in the proportion of total variance explained (3–4% for each additional factor), and similar findings were observed when each of the four disability groups were analyzed separately. The ICLUST algorithm illustrates the relationship between items in terms of Cronbach's alpha and Revelle's beta (worst split-half criterion) for all four groups combined at Time 1 and produced a single cluster (Fig. 1). If desired, multiple clusters could be produced by removing the highest-numbered nodes (e.g., C16 followed by C15 in Fig. 1). Inset in the corner of Fig. 1 is a "scree" plot showing the eigenvalues for the first *N* factors and principal components, suggesting a single-factor model the most parsimonious.

At Time 1, Cronbach's alpha values for the W-ADL were 0.88 in fragile X, 0.90 in autism, 0.91 in Down syndrome, 0.94 in intellectual disability, falling in the upper part of Terwee and colleagues' recommended range (0.7–0.95). The vast majority of inter-item correlations ranged from 0.2 to 0.6 across studies.

Criterion validity

Criterion validity refers to how well the scale actually measures ADLs. Generally, this is tested by correlating the measure to an established "gold standard."¹⁹ However, there is no consensus "gold standard" measure for ADL scales.¹⁸ We correlated ADL scores to the conceptually similar Vineland Screener (Composite Score and "Daily Living Skills" Subdomain Score). Terwee and colleagues recommend

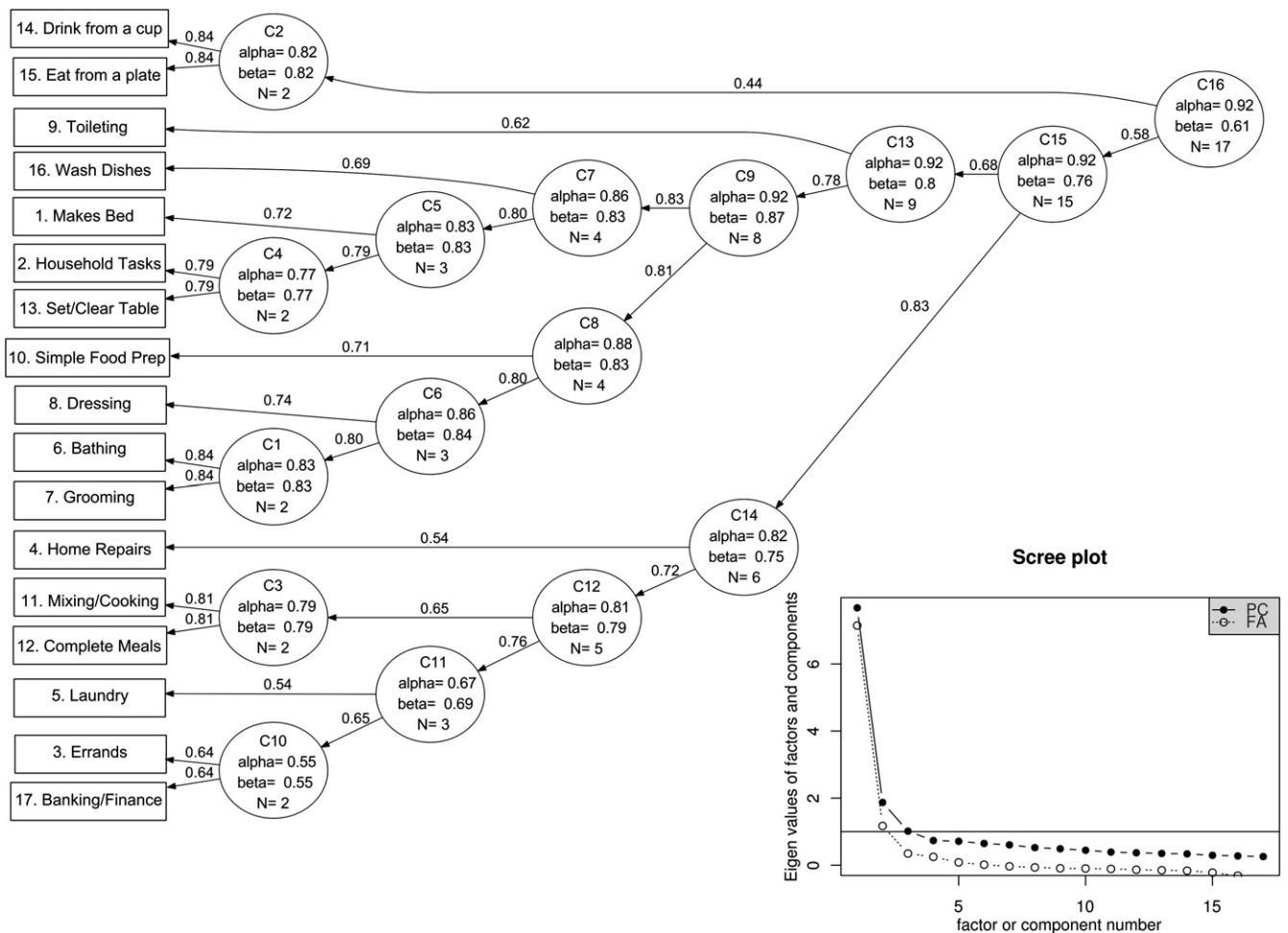


Fig. 1. Hierarchical cluster (ICLUST) analysis of W-ADL items, path values represent reliability-corrected correlation coefficients, as calculated by ICLUST. Scree plot (inset lower right) showing eigenvalues of first N factors and principal components. Both suggest a single-factor structure. Includes all ($N = 989$) participants with complete W-ADL at initial time point.

a correlation of 0.7 or greater with the “gold standard” instrument, which was the standard we applied below.

In the autism sample (at Time 4), the W-ADL measure was highly correlated with the Vineland Screener Adaptive Behavior Composite Score ($r = 0.78$, Fig. 2) and also with the Daily Living Skills subdomain ($r = 0.82$), both greater than Terwee and colleagues’ recommended minimum of 0.7. 49 of the 249 (19.7%) individuals with autism scored the minimum score on the Vineland Daily Living Skills subdomain, suggesting a floor effect on the Vineland Screener. No floor effect was observed on the simultaneously administered W-ADL.

Construct validity

Construct validity refers to the extent to which the W-ADL relates to other characteristics in a manner consistent with the underlying theoretical concept. Prior to analysis, we hypothesized four relationships between W-ADL scores and other measures across the different studies to test the construct validity. We advanced these specific hypotheses based on the

availability of relevant measures and data in each study. Note that the studies used an overlapping—but not an identical—set of measures, so we took advantage of available data to test these hypotheses. We applied Terwee and colleagues’ criterion that 3 of these 4 hypotheses (75%) should be supported by the data and subgroups should be greater than 50 participants.

We made four hypotheses to test the construct validity of the W-ADL. a) We hypothesized that W-ADL scores would be inversely associated with maternally reported caregiving burden (measured on the Zarit Burden Scale) for 147 individuals with fragile X. In a linear model, each W-ADL point was associated with a 0.44 point decrease in caregiving burden ($P < 0.00001$, r -squared = 0.14). b) We hypothesized that W-ADL scores would be positively associated with the probability of participation in supported or competitive employment, or a degree-seeking post-secondary educational program, among 170 adults with autism at Time 4. In a linear model, each W-ADL point was associated with a 2.8% increase in the probability of being employed or in a degree-seeking program ($P < 0.0001$, r -squared = 0.23). c) We hypothesized that

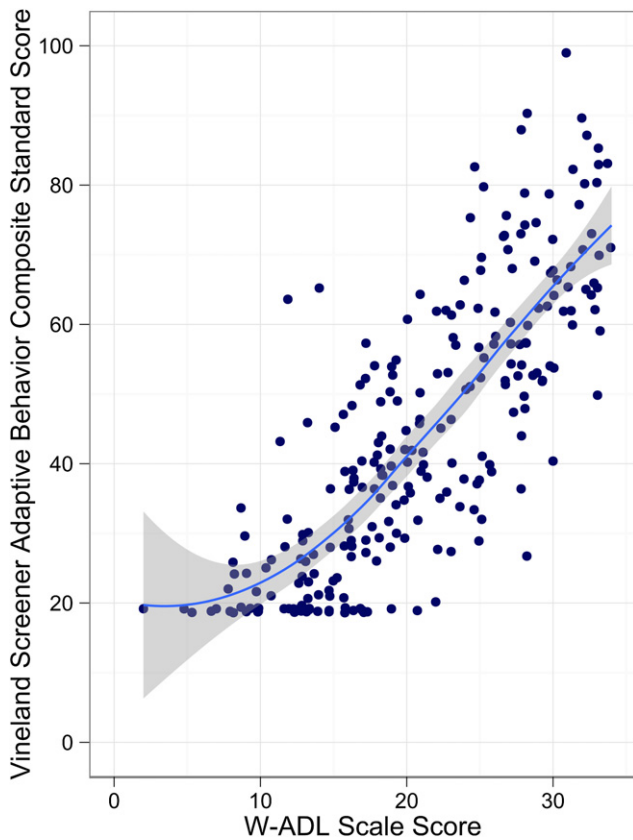


Fig. 2. Scatterplot showing relationship between Vineland Screener adaptive behavior composite score (y-axis) and W-ADL score (x-axis) with loess smoothing line. $N = 249$ adults with autism at Time 4; points slightly jittered to avoid overplotting. Pearson correlation coefficient = 0.78.

W-ADL scores would be positively associated with Stanford-Binet IQ scores among 266 adults with Down syndrome or intellectual disability (all had Stanford Binet IQ ≤ 72). Each W-ADL point was associated with a 0.74 point increase in IQ ($P < 0.00001$, r -squared = 0.15). It is worth noting that 112 of the 266 (42%) scored the minimum of 36 on the Stanford Binet. In addition to the 266, 132 participants were considered untestable or unable to complete the Stanford Binet. d) We hypothesized that W-ADL scores would be inversely associated with maternally reported need or receipt of respite services (within the Down syndrome and intellectual disability study). In a linear model, each W-ADL point was associated with a 2% decrease in reported need or receipt of respite services ($P < 0.00001$, r -squared = 0.12).

Thus, we found support for all four of our a priori hypotheses for relationships between the W-ADL and other measures, exceeding Terwee and colleagues' recommendation that 75% of the hypotheses are supported in subgroups of 50 participants or more.

Reproducibility

Reproducibility is achieved through the minimization of measurement error; a test should show little change in

stable subjects, consistency between raters, and should be capable of discriminating between subjects. To estimate agreement, we calculated the standard error of the measure ($SEM_{\text{agreement}}$) as the square root of the residual error of the intra-class correlation coefficient (ICC) comparing consecutive time points for individuals in the autism and Down syndrome/intellectual disability groups.^{19,32} The $SEM_{\text{individual}}$ was 1.9 in the Down syndrome and intellectual disability groups and 2.0 in the autism group.

Using the $SEM_{\text{agreement}}$, we calculated the smallest detectable change (SDC) for individuals ($SEM \times \sqrt{2} \times 1.96$), and groups ($SDC_{\text{individual}}/\sqrt{\text{participants}}$). Although we report the $SDC_{\text{individual}}$, our primary goal was to evaluate the scale for research use and group comparisons. Therefore, the $SDC_{\text{individual}}$ is 5.2 and 5.5, respectively. Using the larger value of 5.5, the SDC_{group} would fall below 1 point in a study with at least 30 participants.

We measured reliability using a quadratic-weighted kappa (equivalent to the ICC) between consecutive time points for individuals with autism, Down syndrome, or intellectual disability. Among individuals in the Down syndrome and intellectual disability studies, we compared W-ADL scores at Times 1 and 2, and Times 2 and 3. The weighted kappa comparing Times 1 and 2 ($n = 381$) was 0.92, and $\kappa = 0.93$ for Times 2 and 3 ($n = 325$). For adults with autism, Times 7 and 8 (approx. 18 months apart) were the only consecutive time points when the W-ADL was included, and $\kappa = 0.93$ among 209 participants with scores at both times. We also measured the reliability between concurrent mother and father responses for participants in the Down syndrome and intellectual disability groups (the only instance when two raters completed the W-ADL at the same time). Among adults in the Down syndrome and intellectual disability studies at Time 6, the quadratic weighted kappa for inter-rater reliability between 241 complete mother–father responses was 0.88.

These kappa values exceeded the recommended threshold of 0.7, which was the recommended minimum by Terwee and colleagues. The fragile X group only had data from a single measurement available at the time of this manuscript, so reliability could not be calculated for this group.

Responsiveness

Responsiveness refers to the ability to detect changes over time. Although there is little research on the trajectories of ADLs for adults with autism or other developmental disabilities, prior studies have found that, among children with autism, co-occurring intellectual disability was associated with less growth in daily living skills on the Vineland Scales.³³ Thus, we hypothesized that individuals with autism and no intellectual disability would demonstrate greater gains over adolescence and early adulthood, compared to adolescents and adults with autism and intellectual disability. To measure age-related change in these groups, we used mixed-effects regression models with

different parameterizations of age. The addition of either quadratic (age \times age) or inverse (1/age) terms suggested better fit (via the Akaike Information Criterion) over models that only include a linear term for age. We compared these changes over time to the smallest detectable change at the group level.

In the mixed-effects regression models, individuals with autism and no intellectual disability increased an average of 4.9 W-ADL points between the ages of 14 through 25. Individuals with autism and an intellectual disability showed an average increase of just over 1 point (1.1–1.6, depending on parameterization) over the same age range. As the SDC_{group} was well under 1 point in these samples, the W-ADL was able to detect changes over time and to reflect different patterns of change between groups. Smith and colleagues performed a thorough longitudinal analysis confirming the different trajectories by intellectual disability status.³⁴

Floor or ceiling effects

Floor or ceiling effects occur when many participants score at the minimum or maximum values of a measure, and suggest the measure has inadequate range to capture the variability present in a given population. Terwee and colleagues recommend fewer than 15% of participants score the minimum or maximum on a continuous scale, a standard we applied below.

There were no floor or ceiling effects in any of the 4 groups (Down syndrome, intellectual disability, autism, and fragile X) on the W-ADL. Very few participants scored the minimum or maximum level of the W-ADL at any time point; 2 of the 147 participants with fragile X achieved the maximum, and none were at the minimum. Among participants with autism, Down syndrome, or intellectual disability, 1% of participants scored either the maximum or minimum on the scale at the initial time point. This is well within Terwee and colleagues' recommendation that fewer than 15% achieve either extreme score.

Interpretability

Interpretability refers to the ability of the scale to convey qualitative or categorical meaning to the scale score. Terwee and colleagues suggest comparing means and standard deviations for four subgroups.

At the initial time point, data on the "level" of intellectual disability ("mild", "moderate", "severe", "profound") of individuals with Down syndrome or intellectual disability were obtained from agency or school records and confirmed by research administered IQ tests and maternal report. We calculated the mean and standard deviation of W-ADL scores among reported "levels" of intellectual disability at the initial time point and plotted with ggplot2.³⁵ There were significant differences in W-ADL scores (via pairwise *t*-tests, $P < 0.0001$) between all pairs of categories, and the means and standard deviations within each category are presented

in Fig. 3. The smallest group difference in W-ADL scores was 3.1 points, between the "mild" and "moderate" categories; this difference was larger than the smallest detectable group change.

Discussion

This analysis supports the reliability and validity of the W-ADL in research on adolescents and adults with developmental disabilities. By evaluating this measure in the context of three large and established longitudinal studies—that together encompass four major clinical groups (autism, fragile X, Down syndrome, and intellectual disability of no known etiology)—we were able to demonstrate that the W-ADL exceeded Terwee and colleagues' recommended measurement criteria. To our knowledge, this is one of the most comprehensive analyses of the measurement properties of an ADL-based questionnaire.

The 17-item W-ADL appears to be an extremely efficient and practical measure for assessing adults with relatively substantial disabilities. We observed a strong association with the Vineland Screener, which contains 90 items and displayed a floor effect in the autism group. Similarly, the W-ADL was associated with clinician-assessed Stanford Binet IQ scores among the adults with Down syndrome and/or intellectual disability, although many participants scored at the floor of the Stanford Binet or were deemed "untestable". In contrast, the W-ADL was able to detect group differences between subjective or categorical "levels" of intellectual disability, including between severe and profound intellectual disability. Whereas many of the study participants scored at the extreme low end of the

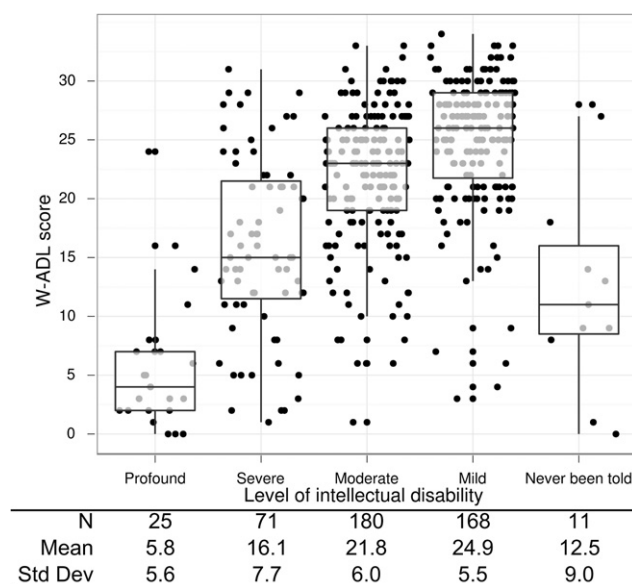


Fig. 3. Boxplot showing W-ADL scores by reported "level" of intellectual disability among 455 adults with Down syndrome and intellectual disability at Time 1. Means and standard deviations for each level shown below figure, showing differentiation between groups.

population distribution on other measures, substantial differences within these groups emerged when using an appropriate and adult-oriented tool.

The majority of the participants in these studies had significant intellectual disabilities or other serious impairments. Therefore, we have limited information related to the level of performance equal to the “ceiling” of the scale. There were 26 participants with autism that had IQs greater than 100 (at Time 4), but the mean W-ADL score for this group was 28.3 and only 1 participant scored the maximum of 34. Future work should consider items that may extend the “ceiling” of the W-ADL. The selection of additional activities should be informed through soliciting suggestions from individuals with the disability (when feasible) and their family members or caregivers. A planned future study to be conducted by our research group will directly interview participants about the determinants of their quality of life, and their responses will help inform the future development of the W-ADL.

This study has some limitations, as the longitudinal studies were not designed for the purpose of evaluating the W-ADL. Although the weighted kappa (or ICC) for consecutive time points was high, real changes in the performance of activities between measurements were interpreted as part of the measurement error, possibly inflating our estimate of the SDC. Increasing the number of items and the range of responses for each item would likely increase the quadratic-weighted kappa and, in turn, decrease the SDC.³⁶ However, these improvements in measurement statistics would likely come at the expense of efficiency (by requiring more items) or the interpretability of the scale units (for instance, if items were scored on a scale from 0 to 5). Another limitation is that our assessment of the W-ADL was based almost exclusively on parent-report approaches. The vast majority of participants were living at home at the start of the study. Future work could examine W-ADL reliability between non-parent respondents.

Like other ADL-based measures, further work is needed to determine the minimum important change in scores.¹⁸ One recommended method for determining the minimum important change is through an anchor-based approach.³⁷ However, because each point on the W-ADL represents partial or complete independence in a specific activity, any change in score could be perceived as important. The smallest important change may depend on the specific activities that change as well the context in which the changes occur. Additional work is also needed to develop inferences that could be made on the basis of W-ADL scores. We found that the W-ADL was associated with perceived caregiving burden, parent-reported need for services, and educational and vocational attainment; these associations could form the basis for a deeper analysis of the W-ADL's predictive and discriminative properties. (e.g., Does a score at a certain age predict future employment or living situation? How well do scores discriminate between people that do and don't qualify for certain services?).

A major strength of this analysis is that it utilizes some of the largest and longest running studies of the phenotype of autism, fragile X, Down syndrome, and intellectual disability in adulthood. In addition to the W-ADL items, numerous other instruments and questionnaires were administered throughout the studies affording the opportunity to examine different aspects of validity. Another important advantage of this study is the inclusion of adults with different categorical diagnoses. While the symptomatology and etiology between fragile X, Down syndrome, autism, and intellectual disability may differ, the W-ADL performed similarly across groups in terms of measurement properties.

An alternative measure to the W-ADL is the Scales of Independent Behavior-Revised (SIB-R) Short Form.³⁸ The SIB-R Short Form is a norm-referenced, 40-item measure of adaptive behavior used to determine a person's level of functioning, and it has been regularly used to assess adults with intellectual or developmental disabilities. Each item is rated on a 4-point scale corresponding to either the proportion of time a person performs—or could perform—a task without help or supervision. Whereas the SIB-R focuses on *capacity* to carry out a task, the W-ADL focuses on the *actual performance* of activities which, as described by the ICF, is concerned with an individual in his or her current environment (which may include assistive devices or other support). The SIB-R Short Form also tends to contain more narrowly defined tasks compared to the W-ADL. For example, the W-ADL contains an item on getting dressed, while the SIB-R Short Form contains multiple items about dressing, including an item on tying shoelaces. Someone wearing shoes with hook-and-loop fasteners (i.e., Velcro) could perform the activity of getting dressed without having the capacity to tie shoelaces, and thus would not be credited in the SIB-R. These conceptual differences could make one measure preferable over the other, depending on whether the goal is to measure specific skills or the lived experience of a person with disability.

The W-ADL may also have utility in disability screening or large surveys, particularly in low-resource settings. Activity- or ADL-based measures have been used in disability screening and their interdisciplinary and cross-cultural relevance offer practical utility in many situations.^{17,39–41} Additional work is needed to assess whether these items are relevant in different contexts and cultures, and to consider how specific activities may be more or less essential to—or representative of—typical life situations. Here, the ICF would serve as a useful framework for identifying important new activities, for understanding relationships between different levels of disability, and for identifying environmental barriers and facilitators for the performance of activities.

In summary, the W-ADL Scale is a freely available (at www.waisman.wisc.edu/family/WADL) measure of activities of daily living for adolescents and adults with developmental disabilities. As a research tool, the W-ADL demonstrated excellent measurement properties across four

well-characterized groups of developmental disabilities including autism, Down syndrome, fragile X, and intellectual disability.

Acknowledgment

We are extremely grateful to the 1014 families who participated in this study; without their generous support and commitment, our research would not be possible.

References

- Durkin MS. Developmental disorders. In: Institute of Medicine (U.S.), ed. *Neurological, Psychiatric, and Developmental Disorders: Meeting the Challenge in the Developing World*. Washington, DC: National Academy; 2001. Print.
- Yeargin-Allsopp M, Murphy CC, Oakley GP, Sikes K, Metropolitan Atlanta Developmental Disabilities Study Staff. A multiple-source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities study. *Pediatrics*. 1992;89:624–630.
- The Developmental Disabilities Assistance and Bill of Rights Act of 2000, PUBLIC LAW 106-402-OCT. 30, 2000 114 STAT. Available at: <http://www.acf.hhs.gov/programs/add/ddact/DDA.html>; 1683 Accessed 08.11.11.
- Brugha TS, McManus S, Bankart J, et al. Epidemiology of autism spectrum disorders in adults in the community in England. *Arch Gen Psychiatry*. 2011;68(5):459.
- Lollar DJ, Crews JE. Redefining the role of public health in disability. *Annu Rev Public Health*. 2003;24:195–208.
- Simeonsson RJ, Lollar D, Hollowell J, Adams M. Revision of the International Classification of Impairments, Disabilities, and Handicaps: developmental issues. *J Clin Epidemiol*. 2000;53(2):113–124.
- Susser M. Disease, illness, sickness; impairment, disability and handicap. *Psychol Med*. 1990;20(03):471–473.
- Nagi SZ. A study in the evaluation of disability and rehabilitation potential: concepts, methods, and procedures. *Am J Public Health Nations Health*. 1964;54(9):1568.
- Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med*. 1994;38(1):1–14.
- World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. World Health Organization; 2001.
- Van Naarden Braun K, Yeargin-Allsopp M, Lollar D. Activity limitations among young adults with developmental disabilities: a population-based follow-up study. *Res Dev Disabil*. 2009;30(1):179–191.
- Rosenbaum P, Stewart D. The World Health Organization International Classification of Functioning, Disability, and Health: a model to guide clinical thinking, practice and research in the field of cerebral palsy. *Semin Pediatr Neurol*. 2004;11:5–10.
- Perry A, Flanagan HE, Dunn Geier J, Freeman NL. Brief report: the Vineland Adaptive Behavior Scales in young children with autism spectrum disorders at different cognitive levels. *J Autism Dev Disord*. 2009;39(7):1066–1078.
- Hessl D, Nguyen DV, Green C, et al. A solution to limitations of cognitive testing in children with intellectual disabilities: the case of fragile X syndrome. *J Neurodev Disord*. 2008;1(1):33–45.
- Couzens D, Cuskelly M, Jobling A. The Stanford Binet fourth edition and its use with individuals with Down syndrome: cautions for clinicians. *Int J Disabil Dev Educ*. 2004;51(1):39–56.
- Mollica RL. *State Medicaid Reimbursement Policies and Practices in Assisted Living*. Prepared for: National Center for Assisted Living, American Health Care Association; 2009.
- Fujiura GT, Rutkowski-Kmitta V. Counting disability. In: Albrecht GL, Seelman KD, Bury M, eds. *Handbook of Disability Studies*. Thousand Oaks, CA: Sage Publications; 2001:69.
- Sikkas SAM, de Lange-de Klerk ESM, Pijnenburg YAL, Scheltens P, Uitdehaag BMJ. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. *J Neurol Neurosurg Psychiatry*. 2009;80:7–12.
- Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
- Seltzer MM, Greenberg JS, Taylor JL, et al. Adolescents and adults with autism spectrum disorders. In: Amaral DG, Dawson G, Geschwind D, eds. *Autism Spectrum Disorders*. New York: Oxford University Press; 2011.
- Lord C, Rutter M, Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*. 1994;24(5):659–685.
- Sparrow SS, Carter AS, Cicchetti DV. *Vineland Screener: Overview, Reliability, Validity, Administration, and Scoring*. New Haven: Yale University Child Study Center; 1993.
- Taylor JL, Seltzer MM. Developing a vocational index for adults with autism spectrum disorders. *J Autism Dev Disord*, in press.
- Smith LE, Barker ET, Seltzer MM, Abbeduto L, Greenberg JS. Behavioral phenotype of fragile X syndrome in adolescence and adulthood. *Am J Intellect Dev Disabil*, in press.
- Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist*. 1980;20:649–655.
- Krauss MW, Seltzer MM. An unanticipated life: the impact of lifelong caregiving. In: Bersani H, ed. *Responding to the Challenge: International Trends and Current Issues in Developmental Disabilities*. Brookline, MA: Brookline Books; 1999.
- Esbensen A, Seltzer MM, Krauss MW. Stability and change in health, functional abilities and behavior problems among adults with and without Down syndrome. *Am J Ment Retard*. 2008;113:263–277.
- Barthel D, Mahoney F. Functional evaluation: the Barthel index. *Md State Med J*. 1965;14:61–65.
- Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9(3 Part 1):179.
- Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois. Available at: <http://personality-project.org/r/psych.manual.pdf>; 2011. Accessed 08.11.11.
- Revelle W. ICLUST: a cluster analytic approach to exploratory and confirmatory scale construction. *Behav Res Methods*. 1978;10(5):739–742.
- de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–1039.
- Baghdadli A, Assouline B, Sonié S, et al. Developmental trajectories of adaptive behaviors from early childhood to adolescence in a cohort of 152 children with autism spectrum disorders. *J Autism Dev Disord*. Available at: <http://www.springerlink.com/index/10.1007/s10803-011-1357-z>; 2011. Accessed 12.10.11.
- Smith LE, Maenner MJ, Seltzer MM. Developmental trajectories in adolescents and adults with autism: the Case of Daily Living Skills. *JAACAP*. 2012;51(6):622–631.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. Available at: <http://had.co.nz/ggplot2/book>; 2009.
- Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology*; 1996:199–202.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371.
- Bruininks RH, Woodcock RW, Weatherman RF, Hill BK. *Scales of Independent Behavior—Revised: Manual*. Boston, MA: Riverside Publishing Company; 1996.
- Cromwell DA, Eagar K, Poulos RG. The performance of instrumental activities of daily living scale in screening for cognitive

- impairment in elderly community residents. *J Clin Epidemiol.* 2003;56(2):131–137.
40. Gottlieb CA, Maenner MJ, Cappa C, Durkin MS. Child disability screening, nutrition, and early learning in 18 countries with low and middle incomes: data from the third round of UNICEF's Multiple Indicator Cluster Survey (2005–06). *Lancet.* 2009;374(9704):1831–1839.
41. Durkin MS, Wang W, Shrout PE, et al. Evaluating a ten questions screen for childhood disability: reliability and internal structure in different cultures. *J Clin Epidemiol.* 1995;48(5):657–666.