The Erlangen Test of Activities of Daily Living: first results on reliability and validity of a short performance test to measure fundamental activities of daily living in dementia patients

Elmar Graessel,[1] Reena Viegas,[2] Renate Stemmer,[3] Brita Küchly,[4] Johannes Kornhuber[5] and Carolin Donath[1]

[1]*Department of Medical Psychology and Medical Sociology, Psychiatric University Clinic Erlangen, Germany*
[2]*Clinic Grabs, Grabs, Switzerland*
[3]*Catholic University of Applied Sciences Mainz, Germany*
[4]*Social Centre "Sophienstrasse" Erlangen, Germany*
[5]*Psychiatric University Clinic Erlangen, Germany*

ABSTRACT

Background: In the absence of an easily applicable performance test for making valid measurements of fundamental activities of daily living (ADL) in dementia patients, this study reports the development of an ADL performance test which constitutes both a reliable and a valid measurement of the relevant autonomous areas of everyday activities for dementia patients.

Methods: The Erlangen Test of Activities of Daily Living (E-ADL-Test) consists of five items: pouring a drink, cutting a piece of bread, opening a small cupboard, washing hands and tying a bow. Each test item underwent standardized evaluation on a scale of 0 to 6. To determine retest reliability each assessment was repeated at two-weekly intervals. The Global Deterioration Scale, Mini-mental State Examination (MMSE) and Nurses' Observations Scale for Geriatric Patients (NOSGER) were used to assess construct validity. Spearman's rank correlation coefficient was applied. Forty-six patients (42 women and 4 men) with clinically diagnosed dementia, who were resident in nursing homes, took part in the validation study. Their average age was 86.

Results: The E-ADL-Test revealed good inter-individual differentiation ability, particularly in cases of moderate to severe dementia. Cronbach's α was 0.77, retest reliability 0.73. The correlation coefficients were -0.47 with GDS, 0.60 with NOSGER and 0.72 with MMSE.

Conclusions: The E-ADL-Test is a suitable performance test for measuring activities of daily living as it is easy to use, reliable, valid and well accepted.

Key words: dementia, activities of daily living, performance test, instrument, reliability, validity

Introduction

Cognitive disorders arising in the course of dementia result in limitations in the patient's ability to carry out activities of daily living (ADL). Intellectually demanding tasks such as telephoning, so-called "instrumental activities of daily living" (IADL) (Lawton and Brody, 1969; Spector *et al.*, 1987; Khin-Heung Chong, 1995), are the first to be affected. As the dementia progresses, fundamental activities of daily living become affected too. The dementia patient becomes increasingly dependent on help from others.

For this reason, limitations of personal independence in carrying out activities of daily living are included in the most frequently used diagnostic guidelines – International Classification of Diseases (ICD-10; World Health Organization, 1999) and *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IVTR; American Psychiatric Association, 2000) – as a central disorder aspect of the dementia syndrome. This aspect is based on the precise psychometric assessment of ability and decline in performance.

IADL and ADL are most frequently documented with the help of ratings performed by other people. The most frequently used instrument

is the Barthel Index (Mahoney and Barthel, 1965). There are further instruments such as the Functional Independence Measure (FIM; IVAR, 1999) the Bayer-ADL (Hindmarch *et al.*, 1998) or, as curtailed measurements, the subscales ADL and IADL of the Nurses' Observations Scale for Geriatric Patients (NOSGER; Spiegel *et al.*, 1991). These measurements can be applied quickly and easily. They presuppose, however, that the assessment is carried out by someone who knows the dementia patient well, and that it reflects the assessor's subjective evaluation.

Few performance tests for measuring the existing ADL ability in dementia patients have been developed to date because they are time consuming. They all have the advantage that they measure the objective degree of performance. The Eurotest (Carnero-Pardo *et al.*, 2006) assesses the patient's ability to deal with money and the DAFA-Test (Direct Assessment of Functional Abilities) addresses activities such as shopping and cooking (Karagiozis *et al.*, 1998). However, the activities measured by both tests are *instrumental* abilities (IADL).

What reasons are there for developing an ADL performance test? The first argument to be mentioned is that IADL performance tests (e.g. in the Eurotest) cannot be performed by many patients with moderate to severe dementia. Moreover, ADL abilities are a decisive factor in maintaining independence in activities of daily living, particularly in a nursing home. The third argument pertains to scientific data collection. In studies on the efficacy of interventions which cannot really be conducted "blind" (e.g. non-medication procedures), measuring ADL abilities "blind" leads to more valid results than the assessment of ADL abilities by another person, as this presupposes extensive information about the patient and the therapies being used, thus increasing the probability of observer bias.

The existing Performance Test of Daily Living (PADL) (Kuriansky and Gurland, 1976), concentrates on basic motor function and does not measure a sequence of activities. The coordination of a sequence of individual activities that the E-ADL-Test requires, however, is an important factor in maintaining independence in everyday activities. For example, self-care is a key dimension of health care. To perform self-care the person has to select and conduct meshing sequences of activities to maintain life, functioning and well-being (Sidani, 2003). For this reason the PADL test fails to cover the ADL range of activities adequately (Skurla *et al.*, 1988).

A valid performance test reflecting the fundamental abilities of daily living has not previously been designed. Therefore it seemed necessary to develop an ADL performance test which constitutes both a reliable and valid measurement that can also be used in nursing homes. The test should differentiate inter-individually, use minimum effort and be well accepted by dementia patients.

Methods

Methodology

The E-ADL-Test is a performance test that has been developed to measure the ability of dementia patients to carry out basic, everyday practical routine tasks related to taking care of themselves, under standardized conditions. The choice and development of the items was carried out in three stages. First, the objective of the test was defined by a group of experts comprising three professional care managers and three scientists. The test should accurately measure the abilities that are closely related to the need for care in both the domestic and institutional environment. The areas of food intake, self-care (including getting dressed and undressed) and autonomy in the living environment were deemed the most important. In the second step, eight tasks that demonstrate these abilities were developed. A preliminary test was carried out on ten dementia patients from another nursing home. In the third step, the appropriate tasks were chosen and the evaluation system was developed. Selection criteria were (i) a test result with the best possible inter-individual differentiation and (ii) minimum effort to enable the tests to be carried out easily. The items "pouring a drink" and "cutting a piece of bread" represent food intake; "washing hands" and "tying a bow" (e.g. shoelaces) represent self-care; and "opening a little cupboard" represents autonomy in the living environment (e.g. ability to open doors or cupboards). These five items demonstrated the best overall and inter-individual differentiation.

Since no established performance test to measure everyday activities in dementia patients previously existed that depicted nursing care relevant to aspects of everyday activities in nursing homes, it was not possible to carry out a study on concurrent validity. The correlation between the E-ADL-Test score and the degree of severity of dementia (Global Deterioration Scale and Mini-mental State Examination score) and the nursing relevant subscales of the NOSGER are useful indications of the construct validity.

The Global Deterioration Scale (GDS) was developed to assess the degree of severity of dementia in elderly people (Reisberg *et al.*, 1982; German version: Ihl and Fröhlich, 1991). The

scale encompasses seven degrees of severity: 1 = no cognitive impairment (not even subjectively), 2 = doubtful cognitive impairment (only subjective impairment), 3 = mild, 4 = moderate, 5 = fairly severe, 6 = severe, and 7 = very severe cognitive impairment. The stages of the GDS correlate strongly with other neuropsychological instruments; retest reliability and inter-rater reliability are reported to be very good (Oremus *et al.*, 2000). The GDS was used to analyze the association with the global level of severity of the dementia syndrome.

In addition, correlation with the finer aspects of cognitive function was ascertained using the Mini-mental State Examination (MMSE; Folstein *et al.*, 1975; German version: Kessler *et al.*, 1990). This is the most frequently used test to screen for dementia and to assess the degree of severity worldwide. The performance test consists of five parts in which 11 cognitive abilities are assessed: spatiotemporal orientation (1, 2), memory span (3), short-term memory (4), attention (5), numeracy (6), language (7), execution of instructions (8), reading (9), writing (10), and constructive praxia (11). The range is scaled from 0 to 30 points. The validity of the MMSE for differentiating between healthy and significantly cognitively disabled persons has been established. However, the measurement has weaknesses in its sensitivity to detecting mild dementia (Ivemeyer and Zerfaß, 2006). The results are not independent of educational level (Schmand *et al.*, 1995) and the discriminatory power of the items varies in different stages of illness (Hampel *et al.*, 2003). Nevertheless, the MMSE is often referred to as the "gold standard" in dementia screening tests.

NOSGER (Spiegel *et al.*, 1991; German version: Spiegel, 1992) uses 30 items to assess the most common characteristics of geriatric patients. Depending on how often a characteristic has occurred in the previous two weeks, each item is rated from 1 to 5 points. Five items are combined to form a subscale, with each subscale – "memory," "IADL," "ADL," "mood," "social behavior," "disturbing behavior" – ranging from 5 points (no disturbance at all) to 25 points (maximum disturbance). Objectivity, reliability, concurrent validity as well as construct validity of the instrument have been reported to be satisfactory (Wahle *et al.*, 1996; Ivemeyer and Zerfaß, 2006).

Statistical procedures

Because of the deviation from the normal distribution in the scores of the E-ADL-Test (tested with the Kolmogorov-Smirnov-Test), Spearman's rank correlation coefficient was used. It was also used for calculation of the retest reliability coefficient. Cronbach's α was used to describe the internal consistency of the E-ADL-Test. Scores from 0.8 upwards are regarded as good (Bortz, 2003) with $p < 0.05$ as the level of significance.

Sample/participants

Dementia patients resident in two nursing homes in Erlangen (Bavaria, Germany) – the social center "Am Ohmplatz" and the "Erlangen Social Centre AWO" – were recruited for the test. At the time of the study, both institutions had sheltered nursing units – "Am Ohmplatz" could accommodate 69 patients and the "Social Centre AWO" 154 (Figure 1). The nursing staff identified residents suspected of having dementia and who fulfilled the inclusion criteria (see Figure 1). Of the 70 residents who were considered suitable candidates for the study, 51 agreed to participate. The ICD-10 criteria of the dementia syndrome (Figure 1) were confirmed in 46 cases by the patients' general practitioners. None of the patients with dementia refused the test or was prevented by an acute illness from participating.

Of the 46 dementia patients, 42 (91%) were female and 4 (9%) male. The average age was 85.9 years (SD 6.6 years). The youngest person examined was 68, the oldest 100. The distribution of the age values did not differ significantly from the normal distribution ($p = 0.767$). The degree of severity of dementia according to the GDS was as follows: GDS = 3, 15 patients (33%); GDS = 4, 10 patients (22%); GDS = 5, 14 patients (30%); GDS = 6, 7 patients (15%). The degree of severity of cognitive impairment, measured using the MMSE, also showed a normal distribution ($p = 0.745$). The MMSE scores ranged from 1 to 28 (mean 16.0; SD 6.1). The test participants showed impairment in all six subscales of the NOSGER. The largest average impairment, 19.0 (SD = 5.7), applied to the "IADL" subscale, followed by "memory" 16.0 (SD = 4.9), "social behavior" 15.4 (SD = 4.3), "mood" 12.2 (SD = 3.4) and "ADL" 11.7 (SD = 4.3).

E-ADL-Test

The test consists of five tasks (items): pouring a drink, cutting a piece of bread, opening a little cupboard, washing one's hands and tying a bow. (For an exact description of the procedure, evaluation and interpretation, see Appendix 1 – available online as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG.)

Exclusion criteria for the E-ADL-Test are severely impaired vision (even when wearing glasses), severely impaired hearing (even when
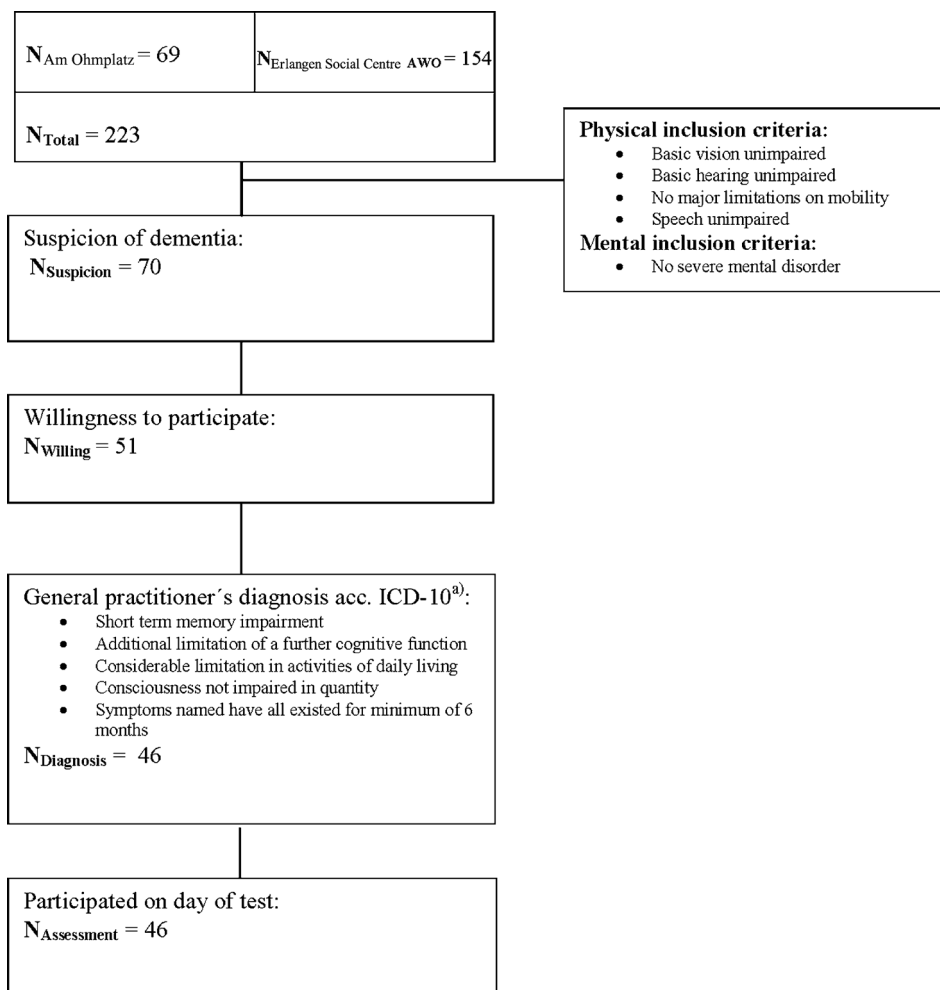
Figure 1. Flow diagram of participant recruitment.

wearing a hearing aid), and severe hand tremor, or paresis of the hand and/or upper extremity which would significantly limit the ability to grip.

The evaluation was based on a points system for each step of a task. The evaluation of each part of the performance test depended on how correctly the activity was performed. Each task is scored on a scale of 0–6 points, 6 points being awarded for a task completed correctly and 0 for a task that could not be performed. The total score is the sum of the scores achieved in the five tasks and ranges from 0 to 30 points, so the higher the score, the better the dementia patient's ability to carry out activities of daily living (ADL-performance). Observing the results of the individual tasks gives an indication of where the patient's deficits lie.

Ethical considerations

The test procedure included a written agreement of participation and was checked and approved by the Ethics Commission at the Medical Faculty of the University of Erlangen-Nuremberg (no. 3232).

Results

Duration of the testing process

The time span for conducting the E-ADL-Test irrespective of preparation or follow-up procedures, i.e. without setting up or removing the test material, ranged from a minimum of 3 minutes to a maximum of 20 minutes. The mean test duration was 8 minutes and 50 seconds ($SD = 4$ minutes 12 seconds). The most common individual time was 5 minutes, which was observed in 13 cases (28%).

Inter-individual variability

The total scores of the E-ADL-Test in the 46 dementia patients show an almost dichotomous distribution of ADL performance (Figure 2), which is skewed left. Ten of the dementia patients (22%)

Table 1. Frequency of test-scores for each item of the E-ADL-Test (first assessment, N = 46)

| E-ADL-TEST ITEM | TASK COMPLETELY SOLVED 6 POINTS | TASK PARTLY SOLVED | | | | | | TASK NOT SOLVED 0 POINTS | TEST SCORE MEAN ± SD[a] | K-S-TEST[b] p[c] |
		5 POINTS	4 POINTS	3 POINTS	2 POINTS	1 POINT	0.5 POINTS			
Pouring a drink	30 (65%)	3 (7%)	2 (4%)	1 (2%)	3 (7%)	5 (11%)	0 (0%)	2 (4%)	4.7 ± 2.1	<0.001
Cutting a piece of bread	17 (37%)	3 (7%)	16 (35%)	4 (9%)	2 (4%)	1 (2%)	0 (0%)	3 (7%)	4.3 ± 1.7	0.031
Opening a little cupboard	21 (46%)	5 (11%)	0 (0%)	3 (7%)	14 (30%)	1 (2%)	0 (0%)	2 (4%)	4.1 ± 2.1	0.002
Washing hands	37 (80%)	1 (2%)	0 (0%)	3 (7%)	0 (0%)	2 (4%)	1 (2%)	2 (4%)	5.2 ± 1.8	<0.001
Tying a bow	30 (65%)	5 (11%)	1 (2%)	1 (2%)	0 (0%)	6 (13%)	0 (0%)	3 (7%)	4.7 ± 2.1	<0.001

[a] mean ± SD: mean ± standard deviation; range: 0 to 6 for each item.
[b] Kolmogorov-Smirnov-Test (K-S-Test).
[c] Level of significance of the K-S-Test: p-values below 0.05 constitute a significant deviation of the normal distribution.
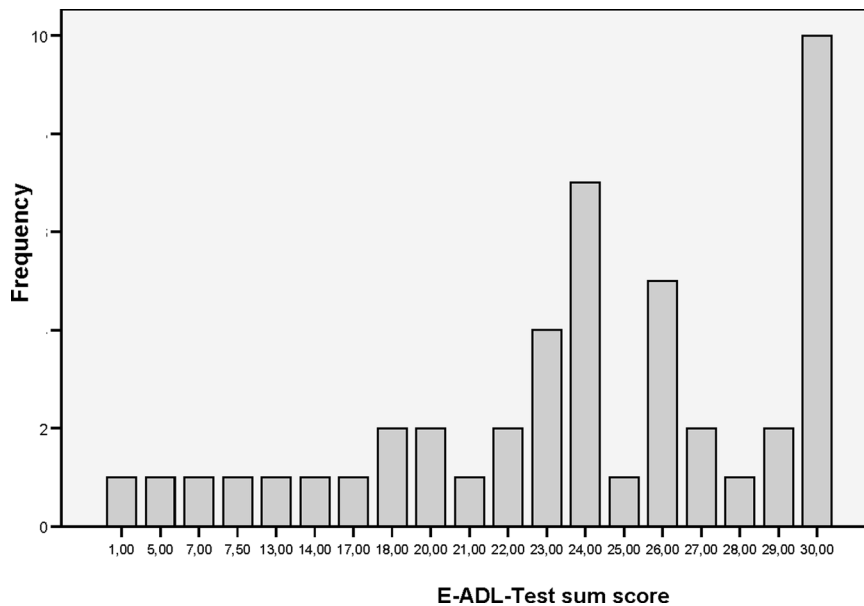
Figure 2. Total scores of E-ADL-Test for the 46 dementia patients.

were able to complete all the tasks correctly (30 points). Among the other 36 participants (78%), the range of possible limitations in performance (0 to 29 points) was almost fully used and ranged in this study from 1 point (not even partial completion of a task) to 29 points (everything right short of one point). For the total sample a moderate deviation from the normal distribution (p = 0.066) can be observed, explained in particular by the 10 cases with 30 points. This resulted in a mean of 23.1 in the total score with a standard deviation of 7.1.

The five tasks resulted in a similar overall picture: the theoretically possible range in performance limitation (0 to 5 points) was observed in all five tasks (Table 1). The percentage of dementia patients with impaired performance ranged from 20% (washing hands), 35% (tying a bow, pouring a drink), 54% (opening a little cupboard) to 63% (cutting a piece of bread).

Reliability

For the E-ADL-Test total score, Cronbach's α was 0.77. The internal consistency of the E-ADL-Test can be described by calculating the Spearman correlation coefficient between the separate items (Table 2). The scores range from 0.18 ("tying a bow" and "washing hands") to 0.51 ("tying a bow" and "pouring a drink"). The median of the correlation coefficients was 0.40.

The test was repeated two weeks later when 42 (91%) of the 46 original participants took part. The

Table 2. Correlation of the items of the E-ADL-Test (first assessment, N = 46)

	POURING A DRINK r_s p	CUTTING A PIECE OF BREAD r_s p	OPENING A LITTLE CUPBOARD r_s p	WASHING HANDS r_s p	TYING A BOW r_s p
Pouring a drink		0.46 *0.001*	0.19 *0.204*	0.33 *0.025*	0.51 *<0.001*
Cutting a piece of bread			0.44 *0.003*	0.41 *0.005*	0.49 *0.001*
Opening a little cupboard				0.30 *0.043*	0.39 *0.008*
Washing hands					0.18 *0.224*
Tying a bow					

r_s: Spearman's rank correlation coefficient.
p: Level of significance of Spearman's rank correlation coefficient.

Table 3. Construct validity of the E-ADL-Test (first assessment, N = 46)

E-ADL-TEST	GDS[a] r_s p	MMSE[b] r_s p	NOSGER: MEMORY[c] r_s p	NOSGER: IADL[d] r_s p	NOSGER: ADL[e] r_s p	NOSGER: MOOD[f] r_s p	NOSGER: DISTURBING BEHAVIOR[g] r_s p	NOSGER: TOTAL SCORE r_s p
Pouring a drink	−0.28 0.062	0.48 0.001	−0.25 0.088	−0.31 0.037	−0.29 0.052	−0.12 0.425	−0.12 0.432	−0.36 0.014
Cutting a piece of bread	−0.43 0.003	0.58 <0.001	−0.34 0.021	−0.43 0.003	−0.22 0.135	−0.24 0.103	−0.31 0.039	−0.48 0.001
Opening a little cupboard	−0.44 0.002	0.63 <0.001	−0.38 0.009	−0.70 <0.001	−0.31 0.038	−0.15 0.318	−0.46 0.001	−0.55 <0.001
Washing hands	−0.04 0.774	0.47 0.001	−0.04 0.807	−0.23 0.129	−0.19 0.201	−0.20 0.194	−0.25 0.093	−0.25 0.097
Tying a bow	−0.39 0.008	0.43 0.003	−0.34 0.020	−0.40 0.006	−0.37 0.011	−0.19 0.217	−0.32 0.032	−0.46 0.001
Total score	−0.47 0.001	0.72 <0.001	−0.37 0.010	−0.57 <0.001	−0.33 0.023	−0.33 0.027	−0.40 0.007	−0.60 <0.001

[a] Global Deterioration Scale: negative correlation with E-ADL-Test, i.e. with increasing GDS-scores (progress in dementia severity) the E-ADL-Test-scores decrease (performance decrease).
[b] Mini-mental State Examination: positive correlation with E-ADL-Test, i.e. with decreasing cognitive performance (decreasing MMSE-scores) the E-ADL-Test-scores also decrease (performance decrease).
[c] Subscale "memory" of the Nurses' Observation Scale for Geriatric Patients, all subscales of the NOSGER correlate negatively with the E-ADL-Test, i.e. with increasing NOSGER-scores (progress in dementia symptoms) the E-ADL-Test-scores decrease (performance decrease).
[d] Subscale "IADL" (instrumental activities of daily living) of the Nurses' Observation Scale for Geriatric Patients.
[e] Subscale "ADL" (activities of daily living) of the Nurses' Observation Scale for Geriatric Patients.
[f] Subscale "mood" of the Nurses' Observation Scale for Geriatric Patients.
[g] Subscale "disturbing behavior" of the Nurses' Observation Scale for Geriatric Patients.
r_s: Spearman's rank correlation coefficient.
p: Level of significance of Spearman's rank correlation coefficient.

reasons why four patients could not be included in the retest were: acute pulmonary infection (one case), hospitalization (two cases) and death (one case). The retest reliability was significant for each item: "pouring a drink" r = 0.35 (p = 0.024), "tying a bow" r = 0.48 (p = 0.001), "cutting a piece of bread" r = 0.56 (p < 0.001), "opening a little cupboard" r = 0.59 (p < 0.001), "washing hands" r = 0.63 (p < 0.001). The retest reliability for the whole test was calculated at 0.73 (p < 0.001).

Validity

The E-ADL-Test total score correlates significantly (p = 0.001) with the severity of the dementia, i.e. the greater the decrease in performance in the E-ADL-Test (lower scores), the greater the degree of dementia (GDS score) or, commensurately, the lower the MMSE score (correlation coefficient $r_s = -0.47$ respectively 0.72; see Table 3). Significant correlations between E-ADL-Test and NOSGER subscales were also found, i.e. a decrease in performance in the E-ADL-Test is reflected in an increase in the NOSGER subscores. This relationship is most evident in the IADL subscale ($r_s = -0.57$; p < 0.001) and least evident in the category "mood" ($r_s = -0.33$; p = 0.027; cf. Table 3).

E-ADL-Test performance and severity of dementia

Categorizing the degree of severity of dementia into mild, moderate and severe according to the MMSE shows that only patients with mild or moderate dementia are able to complete the five E-ADL-Test tasks without making any mistakes, thus obtaining the maximum score of 30 (Table 4). According to the results of the validation sample presented here, there is a strong likelihood (80%) that these people have mild dementia. There were only a few participants with moderate dementia who were able to complete the E-ADL-Test without making any mistakes (2 out of 21).

The E-ADL-Test was completed without making any mistakes by 42% of the 19 patients with mild dementia but 58% of them showed slight performance impairment (29 to 23 points). The E-ADL-Test differentiated well between moderate and severe dementia (see Table 4). An overlap only occurred in one out of 27 patients (8 points in the MMSE and 18 points in the E-ADL-Test). The remaining cases of severe dementia all had E-ADL-Test scores under 14 and thus were not classified in the moderate dementia category of 14 to 30 E-ADL-Test points.

Discussion

The E-ADL-Test was developed because no ADL test was available to measure ability in activities of daily living in a way that was practicable, simple, reliable and valid. Both its immediate relevance to everyday matters and the relatively small number of five items involved make this instrument easy to use in both the nursing home and the wider community. Since it takes only 8 minutes on average to conduct, the E-ADL-Test is not overtaxing for the participants and can be easily learned by nursing staff. Because of the everyday nature of the tasks, they can easily be integrated in the nurses' workflow.

The validation study showed that its acceptance was very high. All 42 dementia patients who were available for the retest took part in it. The advantages of the instrument in comparison to already existing measures can be summarized as follows: (1) The performance test assesses a sequence of actions relevant to care in nursing homes and in the community. (2) The test can be conducted very easily. (3) The instrument enables inter-individual differentiation especially in patients with moderate or severe dementia. (4) The instrument allows objective measurement of ADL-competence and is therefore suitable for evaluation trials of pharmacological and non-pharmacological therapies for dementia patients.

The retest reliability of 0.73 after two weeks indicates that the test is well suited to longitudinal assessment; its degree of sensitivity for measuring symptom changes needs to be researched in future studies. The internal consistency as a measure of reliability lies very close to 0.8, indicating that the procedure is reliable.

As expected, there is a close relationship ($r_s = 0.72$) between the E-ADL-Test result and the patient's cognitive impairment as measured by the MMSE. This is to be expected inasmuch as understanding and executing the test instructions depend on the level of the patient's cognitive skills. The clearly smaller correlation in the cases of mild dementia shows, however, that cognition and activities of daily living correlate but are, in fact, different concepts.

The correlation with the NOSGER subscales varied. The convergent validity of the E-ADL-Test score can be seen in the fact that the strongest correlation was with the "IADL" factor ($r_s = -0.57$), whereas the weakest was with the "mood" factor ($r_s = -0.33$) demonstrating discriminant validity.

In the process of developing the E-ADL-Test, care was taken to choose tasks which were closely related to activities of daily living. The single tasks in the E-ADL-Test pick up on sequences of actions

Table 4. E-ADL-Test-performance depending on dementia severity (first assessment, N = 46)

MMSE	DEMENTIA SEVERITY	N[a]	E-ADL-TEST		MEAN ±SD[d]
			MIN.[b]	MAX.[c]	
0–9	severe	6	1	18	8.6 ± 6.0
10–17	moderate	21	14	30	23.0 ± 3.9
>= 18	mild	19	23	30	27.7 ± 2.6

[a] Number of cases.
[b] lowest value.
[c] highest value.
[d] mean ± standard deviation.

that are likely to occur frequently in the daily life of a nursing home resident. The test participants' basic familiarity with these tasks is shown by the degree to which the participants were able to carry out these tasks depending on the severity of their dementia (Table 4). Remarkably good E-ADL-Test scores were achieved even in cases of moderate dementia.

There are no general limitations in applying the test since the basic abilities required to carry out the tasks, such as pouring a drink, are learned independently of any culturally specific features. However, in applying the test, individual exclusion criteria must be borne in mind. Severely depressed patients, for example, may not be motivated to perform the various actions. Similarly, because this test is limited to the functioning of the upper extremities (tests already exist for the functioning of the lower extremities, e.g. the timed "Up and Go" test; Podsiadlo and Richardson, 1991), those with paralysis of the upper extremities are also excluded. Further research is required to clarify the extent to which the tasks examined in the E-ADL-Test permit conclusions to be drawn about performance in other categories of activities of daily living. In the next step, a reference sample that is representative of patients of both genders and covering all degrees of dementia needs to be recruited, and secondary test criteria such as inter-rater reliability need to be set down.

Conflict of interest

None.

Description of authors' roles

E. Graessel designed and supervised the study, provided statistical expertise for the data analysis and revised the manuscript. R. Viegas collected and analyzed the data and drafted parts of the manuscript. B. Küchly collected data, provided administrative and technical support and revised the manuscript. R. Stemmer participated in the study

conception and drafted parts of the manuscript. J. Kornhuber supported the data analysis, was part of the conception team behind the study and revised the manuscript. C. Donath carried out data analysis, drafting of the manuscript and provided statistical expertise.

Supplementary material
Appendix 1 provides an exact description of the items contained in the instrument, and the procedure, evaluation and interpretation of the test. Appendix 2 is the test form as presented to the users of the E-ADL-Test.

Both appendices are published as supplementary material online attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, DSM-IVTR* (4th edition, text revision). Washington DC: American Psychiatric Association.

Borst, J. (2003). *Statistik für Sozialwissenschaftler* [*Statistics for Social Scientists*], Berlin v.a.: Springer.

Carnero-Pardo, C. *et al.* (2006). Diagnostic accuracy of the Eurotest for dementia: a naturalistic, multicenter phase II study. *BMC Neurology*, 6, 15.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Hampel, H., Padberg, F. and Möller, H.-J. (2003). *Alzheimer-Demenz* [Alzheimer's Disease], Stuttgart: Wissenschaftliche Verlagsgesellschaft.

Hindmarch, I., Lehfeld, H., Jongh, P. and Erzigkeit, H. (1998). The Bayer Activities of Daily Living Scale (B-ADL). *Dementia and Geriatric Cognitive Disorders*, 9 (Suppl. 2), 20–26.

Ihl, R. and Fröhlich, L. (1991). *Die Reisberg-Skalen GDS, CRS, FAST* [The Reisberg Scales GDS, CRS, FAST], Weinheim: Beltz Test.

IVAR (Internationale Vereinigung für Assessment in der Rehabilitation) [International Society for Assessment in Rehabilitation] (1999). *Manual FIM – Funktionale Selbständigkeitsmessung* [FIM – Functional Independent Measure], Straubing: IVAR.

Ivemeyer, D. and Zerfaß, R. (2006). *Demenztests in der Praxis – ein Wegweiser* [Dementia Tests in Practice – A Guide], München, Jena: Urban and Fischer.

Karagiozis, H., Gray, S., Sacco, J., Shapiro, M. and Kawas, C. (1998). The Direct Assessment of Functional Abilities (DAFA): a comparison to an indirect measure of Instrumental Activities of Daily Living. *The Gerontologist*, 38, 113–121.

Kessler, J., Markowitsch, H. J. and Denzler, P. (1990). *Mini-Mental-Status-Test*. Weinheim: Beltz Test.

Khin-Heung Chong, D. (1995). Measurement of instrumental activities of daily living in stroke. *Stroke*, 26, 1119–1122.

Kuriansky, J. and Gurland, B. (1976). Self care capacity: the performance tests of activities of daily living. *International Journal of Aging and Human Development*, 7, 343–352.

Lawton, M. P. and Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist*, 9, 179–186.

Mahoney, F. I. and Barthel, D. W. (1965). Functional evaluation: the Barthel Index. *Maryland State Medical Journal*, 14, 61–65.

Oremus, M., Perrault, A., Demers, L. and Wolfson, C. (2000). Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of global scales. *Journal of Geriatric Psychiatry and Neurology*, 13, 197–205.

Podsiadlo, D. and Richardson, S. (1991). The timed "Up and Go": a test of functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39, 142–148.

Reisberg, B., Ferris, S. H., Leon, M. J. de and Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139, 1136–1139.

Schmand, B., Lindenboom, J., Hooijer, C. and Jonker, C. (1995). Relation between education and dementia: the role of test bias revisited. *Journal of Neurology, Neurosurgery, and Psychiatry*, 59, 170–174.

Sidani, S. (2003) Self-care. In D. Dorane (ed.) *Nursing-sensitive Outcomes* (pp. 65–113). Boston: Jones and Bartlett.

Skurla, E., Rogers, J. C. and Sunderland, T. (1988). Direct assessment of activities of daily living in Alzheimer's disease: a controlled study. *Journal of the American Geriatrics Society*, 36, 97–103.

Spector, W. D., Katz, S., Murphy, J. B. and Fulton, J. P. (1987). The hierarchical relationship between activities of daily living and instrumental activities of daily living. *Journal of Chronic Diseases*, 40, 481–489.

Spiegel, R. (1992). Erfassung des Verhaltens psychogeriatrischer Patienten im Alltag mit der NOSGER [Recording the behaviour of psycho-geriatric patients in daily life using NOSGER]. In N. I. Jovic and A. Uchtenhagen (eds.), *Ambulante Psychogeriatrie: neue Wege und Hinweise für die Praxis* [Outpatient Psychogeriatrics: New Ways and Tips]. Kröning: Roland Asanger Verlag.

Spiegel, R. *et al.* (1991). A new behavioral assessment scale for geriatric out- and inpatients: the NOSGER (Nurses' Observation Scale for Geriatric Patients). *Journal of the American Geriatrics Society*, 39, 339–347.

Wahle, M., Häller, S. and Spiegel, R. (1996). Validation of the NOSGER: reliability and validity of a caregiver rating instrument. *International Psychogeriatrics*, 8, 525–547.

World Health Organization (1999). *International Classification of Diseases (ICD-10)*. *Chapter V: Mental and Behavioral Disorders* (Trans. H. Dilling, W. Mombour and M. H. Schmidt) Bern: Huber.