




Normative Data for the CERAD-NP for Healthy High-Agers (80–84 years) and Effects of Age-Typical Visual Impairment and Hearing Loss

Stephanie Fröhlich^{1,2} , Katrin Müller^{2,3}  and Claudia Voelcker-Rehage^{1,2,*} 

¹Department of Neuromotor Behavior and Exercise, Institute of Sport and Exercise Sciences, Faculty of Psychology & Sport Sciences, University of Münster, Münster, Germany

²Department of Sports Psychology (With Focus on Prevention and Rehabilitation), Institute of Human Movement Science and Health, Faculty of Behavioural and Social Sciences, Chemnitz University of Technology, Chemnitz, Germany

³Department of Social Science of Physical Activity and Health, Institute of Human Movement Science and Health, Faculty of Behavioural and Social Sciences, Chemnitz University of Technology, Chemnitz, Germany

(RECEIVED February 18, 2021; FINAL REVISION October 12, 2021; ACCEPTED October 25, 2021)

Abstract

Objectives: This study aims to establish reference data for nondemented adults between 80 and 84 years of age based on the German version of the Consortium to Establish a Registry for Alzheimer’s disease Neuropsychological (CERAD-NP) test battery and to assess the possible influence of hearing and vision impairments on CERAD-NP performance. **Methods:** Two hundred one volunteers were examined with the German CERAD-NP test battery, and 18 test scores were calculated from the data. The sample included 99 men (49%), the mean age was 81.8 years ($SD = 1.3$), and the mean years of education were 13.9 ($SD = 3.1$). Percentiles for continuous and percentile ranks for discrete test scores were calculated separately for four norm groups. The groups were classified according to gender and education. Multiple regression analysis was used to predict cognitive performance from visual acuity and hearing ability. **Results:** The normative data obtained were consistent with other findings from younger and older age groups. Worse visual acuity predicted slower performance in the Trail Making Test (TMT). None of the other CERAD-NP tests were correlated to sensory functions. **Conclusions:** Using age-appropriate reference data, such as that established here for the 80–84 year age group can help to improve the detection of cognitive decline and prevent biases that arise when old-old adults are compared to younger old adults. Visual acuity should be considered an influencing factor on TMT performance.

Keywords: Aging, neuropsychological tests, neurocognitive disorders, mild cognitive impairment, reference values, sensory impairment

Neuropsychological test batteries are commonly used in clinical settings and research to detect cognitive decline and the presence of dementia in older adults (OA). One of the most established test batteries was developed by the Consortium to Establish a Registry for Alzheimer’s disease (CERAD, Morris et al., 1989) and has since been revised, expanded, and translated into more than 20 different languages (Duke Aging Center, n.d.; Fillenbaum et al., 2008). The CERAD-Neuropsychological (CERAD-NP) test battery was developed to measure cognitive deficits in patients with Alzheimer’s disease of different stages

including the early stages (Morris et al., 1989). The CERAD-NP has good validity, reliability, and high acceptance among health professionals (Fillenbaum et al., 2008; Morris et al., 1993). It has also been shown to be a suitable tool for detecting mild cognitive impairment (Breton, Casey, & Arnaoutoglou, 2019).

The detection of mild impairment requires normative data against which individual test scores can be evaluated so that it can be determined whether they differ notably from the performance expected from a healthy individual. The first comprehensive normative data set of healthy OA (50–89 years) for the CERAD-NP clearly showed the influence of age, sex, and education on test performance (Welsh et al., 1994). In addition, language and cultural background are also considered to be influencing factors

*Correspondence and reprint requests to: Claudia Voelcker-Rehage, Department of Neuromotor Behavior and Exercise, Institute of Sport and Exercise Sciences, University of Münster, 48149 Münster, Germany. E-mail: claudia.voelcker-rehage@uni-muenster.de

(Paajanen et al., 2010). By now many additional normative data sets have been published (i.e., Kirsebom et al., 2019; Liu et al., 2011; Luck et al., 2018; O'Bryant et al., 2018).

While most of these data sets span a wide age range, the usefulness for OA beyond the age of 80 is limited because normative data sets for high-agers are often small and not well balanced according to gender or education (Miller et al., 2015). A few studies have tried to combat this bias by studying the CERAD-NP performance of high-agers specifically. Beeri and colleagues (2006) obtained normative data from a sample of 196 healthy individuals 85 years and over in the USA. Additionally, Luck and colleagues (2009) published data for the age group >75 years in Germany, but only included the memory subtests of the CERAD-NP. Both studies concluded that using norms based on younger cohorts or only small, biased samples of an older age group can lead to more false positives and the subjective interpretation of results (Beeri et al., 2006; Luck et al., 2009).

Especially in the German-speaking area, there is a lack of normative data in this age range (>80 years) beyond the memory subtests. The German version of the CERAD-NP (Memory Clinic Basel (2005)) includes two additional tests (Trail Making Test and Verbal Fluency with S-Words) that measure executive functions (Schmid, Ehrensperger, Berres, Beck, & Monsch, 2014). This version was validated (Aebi, 2002), and normative data are available for the original CERAD-NP tests based on a sample of 1100 Swiss healthy OA (49–92 years). The current normative data for the additional CERAD-NP tests are based on 604 Swiss healthy OA (50–88 years). However, the sample sizes of both these data sets are unequally distributed across age groups. For example, only one person was included for highly educated females in the age group >80 years. Luck et al. (2018) published an update of the CERAD-NP norms based on data obtained between 2011 and 2014, but only included the age range from 60 to 79 years.

This presents a good basis for applying the CERAD-NP to older samples both in Germany and internationally. Considering the increasing number of individuals older than 80 years in western societies and the concomitant increase of patients with dementia (Deutsche Alzheimer Gesellschaft e.V., 2018; Statistisches Bundesamt (Destatis), 2019), the frequency of and need for neuropsychological testing in this age group will continue to grow. Hence, reliable and comprehensive data sets for neuropsychological tests are required. Furthermore, normative data must be updated regularly to account for cohort effects and socio-environmental changes (Dickinson & Hiscock, 2011). Thus, the primary objective of this study is to complement already published data by providing a current and comprehensive normative data set for individuals between 80 and 84 years of age that is based on all CERAD-NP subtests. Our sample can be considered part of the old-old age group, which is used as a term to describe adults between 75 and 85 years old (Boyd & Bee, 2006). As a secondary outcome, we use the same sample

to investigate how performance relates to sight and hearing performance.

INFLUENCE OF HEARING AND VISION

Standard cognitive tests almost exclusively use visual and auditory stimuli and oral test instructions. Furthermore, visual impairments like cataracts, glaucoma or macular degeneration, and hearing loss are very common in populations over 80 years of age (Hesse, Eichhorn, & Laubert, 2014; Reitmeir et al., 2017). Even though treatment and support with aids are routinely available, there is still a portion of individuals that do not regularly use their aids (Ober, Marcusson, Nagga, & Wressle, 2012; Tsai, 2009) or who have impairments that cannot be sufficiently corrected or reversed by treatment (Nowak, 2006). Considering this, epidemiological study samples, as well as individuals receiving neuropsychological testing, probably cover a broad range of sensory functioning, and this must be considered when testing them.

As expected, impairments in visual and hearing ability have been shown to result in poorer performance in the Mini-Mental Status Examination (MMSE), part of the CERAD-NP, and other screening tools (Dupuis et al., 2015; Lim & Loo, 2018). To date, the relationship between hearing or vision and CERAD-NP performance in OA has not been clarified in any known studies. This knowledge is needed to better interpret the performance of OA with sensory impairments. Therefore, the secondary aim of this study is to examine the association between corrected visual acuity and corrected hearing ability and CERAD-NP performance.

Cross-sectional and longitudinal data have suggested that a substantial amount of variance in cognition can be explained by the quality of sensory functions (Li & Lindenberger, 2002). This might be because there are reduced resources available for cognitive tasks as additional cognitive effort is required for perceptual success in the presence of sensory deficits (McCoy et al., 2005; Wood et al., 2010), the so-called effortfulness hypothesis. Alternatively, the common cause hypothesis assumes that common neuropathological processes account for the changes in sensory and cognitive function (Uchida et al., 2019). A mixture of both explanations is most likely responsible for the strong connections between sensory and cognitive functioning during old age (Li & Lindenberger, 2002).

METHODS

Sample

The nondemented volunteers were recruited as part of the SENDA (Sensor-based systems for early detection of dementia) study at the Chemnitz University of Technology, Germany. The study was approved by the Ethics Committee of the Chemnitz University of Technology (Faculty of Behavioral and Social Sciences) on December

Table 1. Recruitment strategies, inclusion, and exclusion criteria of the SENDA study

Recruitment strategies
<ul style="list-style-type: none"> – calls for participation via (free) local newspapers – calls for participation via the official university website – invitation letters sent to 3300 Chemnitz residents in cooperation with the registration office (random selection from addresses with the following criteria: German citizens, age 80–90 years, no nursing homes) – word of mouth from already enrolled volunteers
Inclusion criteria
<ul style="list-style-type: none"> – Age ≥ 80 years^a – Independent means of travel to and from the testing facility – German fluency at native language level
Exclusion criteria
<ul style="list-style-type: none"> – Medical ban from sports and other strenuous activities – Diagnosed psychological disorders (e.g., major depressive episode, anxiety disorder, substance use disorder) – Diagnosed neurocognitive disorders (e.g., delirium, dementia due to Alzheimer’s disease, dementia due to vascular disease) – Montreal Cognitive Assessment < 19 – Permanent impairments due to brain surgery or stroke – Other neurological diseases (e.g., epilepsy, Parkinson’s disease, neuropathy) – Severe diseases of the respiratory system (e.g., COPD stage 4, severe asthma) – Severe diseases of the cardiovascular system (e.g., cardiac arrhythmia, heart failure, arterial occlusive disease) – Severe diseases of the musculoskeletal system (e.g., severe arthritis, orthopedic operations in the last 6 months) – Diabetes with diagnosed neuropathy – Substance abuse – Current participation in other clinical trials

^aParticipants were also included if they turned 80 during the course of the baseline measurements which included 3 separate testing days.

19, 2017 (V-232–17-KM-SENDA-07112017) and is included on the German Clinical Trials Register (DRKS00013167). Recruitment strategies as well as inclusion and exclusion criteria can be found in Table 1. Among others the following exclusion criteria were applied here: diagnosed psychological disorders (e.g., major depressive episode, anxiety disorder, substance use disorder) and diagnosed neurocognitive disorders (e.g., delirium, dementia due to Alzheimer’s disease, dementia due to vascular disease). Eligibility was determined via telephone interview carried out by a trained study nurse. Furthermore, the face-to-face Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005) and the MMSE (Folstein, Folstein, & McHugh, 1975) as part of the CERAD-NP were carried out.

For more information refer to the SENDA study protocol by Müller et al. (2020).

Between January 2018 and March 2020, 201 volunteers (born between 1933 and 1939, age 80–84 years, $M = 81.8$, $SD = 1.3$) were recruited in Chemnitz and its surroundings. This five-year age range was chosen to ensure comparability

with other normative data sets (Beeri et al., 2006; Luck et al., 2018) and to prevent biases arising from wide age ranges (Miller et al., 2015). Neither the younger (79 years, $n = 8$) nor the older (85–91 years, $n = 35$) participants from the SENDA study were used here because the numbers were deemed too small to be representative. The sample was well balanced according to gender (99 males and 102 females) and included 122 highly educated (>12 years of education) compared to 79 less educated individuals (≤ 12 years of education). The corrected hearing and visual acuity status of participants were representative of independently living old-old adults. This sample incorporated impairments ranging from normal functioning to moderate, but excluded impairments that would inhibit independent living or activities of daily living. Table 2 contains additional sample characteristics.

Material

Sociodemographic variables

The sociodemographic variables age, gender, and years of education (including school and further professional education) were obtained via a short, structured interview prior to neuropsychological testing. Education was dichotomized into high level of education (>12 years of education) and low level of education (≤ 12 years of education) according to Welsh et al.’s (1994) classification system.

CERAD-NP

The extended CERAD-NP was carried out by trained staff and strictly followed the manual provided by the Memory Clinic Basel. This included the following tests: Verbal Fluency Animals and S-Words, Boston Naming Test, MMSE, Wordlist Learning, Recall and Recognition, Constructional Praxis Copying and Recall, Trail Making Test (TMT) A and B. Eighteen test scores were calculated from these tests (Table 3). The only change pertained to the presentation of visual stimuli for the Boston Naming Test (pictures) and the Wordlist Learning and Recognition (words). A custom-made LabView 2015 (National Instruments, Austin, TX, USA) script was used to present stimuli in the center of a screen (using the same size and font as the original stimuli) for standardized implementation.

Sensory testing

During testing, participants used the same aids (i.e., glasses and/or hearing devices) they normally use during everyday life. Corrected visual acuity was determined by the Freiburg Visual Acuity Test with Landolt C (Bach, 1996). Participants sat three meters from the screen and completed 18 trials to obtain the logarithm of the minimum angle of resolution (logMAR). This parameter is a measure of visual acuity loss and logMAR scores from 0 to .5 are considered (near

Table 2. Sample characteristics for the total sample and subgroups according to gender and years education and results for ANOVAs with factor group

	Total (<i>N</i> = 201)	Male (<i>N</i> = 99)		Female (<i>N</i> = 102)		<i>F</i> statistic			
		>12 years education (<i>N</i> = 75)	≤12 years education (<i>N</i> = 24)	>12 years education (<i>N</i> = 47)	≤12 years education (<i>N</i> = 55)	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2
Demographics									
Age (M (SD))	81.8 (1.3)	81.7 (1.3)	82.2 (1.2)	81.4 (1.4)	82.0 (1.3)	2.4	3,200	.07	.09
Education years (M (SD))	13.9 (3.1)	16.6 (2.5)	10.7 (1.1)	14.7 (1.7)	10.9 (0.8)	129.1	3,200	<.001 ^{a,b,c,d,e}	.66
Global Cognition									
MMSE (M (SD))	27.9 (1.6)	28.1 (1.3)	27.3 (1.8)	28.6 (1.5)	27.3 (1.7)	7.4	3,200	<.001 ^{b,d,e}	.18
MoCA (M (SD))	25.6 (2.6)	25.4 (2.6)	24.7 (2.6)	26.0 (2.5)	25.8 (2.8)	1.8	3,200	.15	.07
Health*									
CCI (M (SD))	0.8 (1.3)	0.7 (1.2)	1.4 (1.5)	0.8 (1.5)	0.8 (1.0)	1.8	3,185	.15	.08
GDS (M (SD))	3.2 (2.5)	2.5 (1.9)	4.2 (2.9)	2.8 (2.3)	4.0 (3.0)	5.7	3,190	.001 ^{a,d}	.15
SWL (M (SD))	5.5 (1.0)	5.7(0.8)	5.7 (1.0)	5.3 (1.0)	5.3 (1.1)	3.3	3,191	.02	.11
Cardiovascular diseases (%)	67	64	70	68	70				
Respiratory diseases (%)	16	7	35	14	22				
Diabetes (%)	15	13	9	11	24				

Note. Superscripts a-e denote significant ($p < .05$) differences between: Male > 12 years and Male ≤ 12 years (a), Females > 12 years and Females ≤ 12 years (b), Males > 12 years and Females > 12 years (c), Males > 12 years and Females ≤ 12 years (d), Males ≤ 12 years and Females > 12 years (e). MMSE = Mini-Mental Status Examination (Folstein et al. 1975), MoCA = Montreal Cognitive Assessment (Nasreddine et al., 2005), GDS = Geriatric Depression Score (15-item version, Gauggel & Birkner, 1999), SWL = Satisfaction with Life Scale (mean score, Diener, Emmons, Larsen, & Griffin (1985), CCI = Charlson Comorbidity Index (Charlson, Pompei, Ales & MacKenzie, 1987).

*Health data was only available for $n = 189$ and all measures were self-reports.

normal vision, while higher scores are classified as low vision (6–13) or blindness (≥ 14) (Colenbrander, 2002).

To quantify corrected hearing performance, one practice list (18) and three test lists (4, 14, 20) from the Freiburg monosyllabic test (part of the Freiburg speech test (Hahlbrock, 1953)) were presented at four sound levels (35 dB, 47 dB, 24 dB, 53 dB) without background noise via headphones (SHARK ZONE H10 Gaming Stereo-Headset, Sharkoon Technologies GmbH, Germany). The same order was used for all participants and the number of correctly repeated words (out of 20) was recorded for each test list. The rate of understanding at the 24 dB sound level was calculated as a percentage because this list displayed the widest range (0–20) and greatest variance ($SD = 5.17$) of the test lists.

Statistical Analysis

The analysis was done with SPSS IBM Statistics Version 27 (IBM Corp., Armonk, NY, USA). For each CERAD-NP score, a 2×2 analysis of variance (ANOVA) with between-subject factors sex (male/female) and education (high/low) was used to determine whether normative data should be calculated for the whole sample or subdivided into different groups. The results indicated that only eight scores were not significantly influenced by either gender or level of education (Table 4). Therefore, all further analyses were done

separately for the following groups: (1) males with >12 years of education, (2) males with ≤12 years of education, (3) females with >12 years of education, and (4) females with $12 \leq$ years of education. Mean, standard deviation, minimum, maximum, skew, and kurtosis were calculated for each score, and distributions were tested for normality with Shapiro-Wilk tests.

Percentile ranks (PR) for discrete test scores and percentiles (2.28, 6.68, 10, 15.87, 25, 50, 75, 90) for continuous test scores were calculated because the majority of variables were not normally distributed and therefore did not allow for the calculation of standard norms. Afterward, standard norm equivalents in the form of z-scores were calculated using area transformation (Lienert & Raatz, 1998). The detailed steps are explained in the supplement. PR are only ordinal scales but can be easily interpreted for individual diagnostics, because they show how common an individual's test score is (Crawford, Garthwaite, & Slick, 2009). Z-scores are interval scales that can be used for group statistics and the interpretation of differences (Woerner, Müller, & Hasselhorn, 2017). In addition, they can be transformed into all other commonly used scales such as T or IQ scales by linear transformation.

In the final phase of analysis, multiple linear regression analyses for the whole sample were carried out with predictors age, gender, and years of education in a first step to control for these demographic variables. Visual acuity and

Table 3. List of all CERAD-NP subtests including the name, description, and range of all outcome scores in order of testing (based on information from the German CERAD-NP manual (Memory Clinic Basel, 2005))

Test	Score	Description	Range
Verbal Fluency Animals	Fluency Animals	Number of correct answers	0–
Boston Naming Test	Boston Naming	Number of correct answers	0–15
MMSE	MMSE	Sum score of all MMSE items	0–30
Wordlist Learning	List 1	Number of correct remembered words of list 1	0–10
	List 2	Number of correct remembered words of list 2	0–10
	List 3	Number of correct remembered words of list 3	0–10
	Total	Number of correct remembered words of all three lists	0–30
Constructional Praxis	Copying	Number of criteria fulfilled for all copied figures	0–11
Wordlist Recall	Recall	Number of correct recalled words	0–10
	Savings ^a	Proportion (in %) of correctly recalled words calculated as $\frac{Recall * 100}{List\ 3}$	0–
	Intrusions ^b	Overall number of intrusions during Learning (List 1–3) and Recall	0–
Wordlist Recognition	Recognition ^a	Proportion (in %) of correctly classified words	0–100
Constructional Praxis	Recall	Number of criteria fulfilled for all recalled figures	0–11
	Savings ^a	Proportion (in %) of criteria fulfilled for all recalled figures calculated as $\frac{Recall * 100}{Copying}$	0–
Trail Making Test	TMT A ^{a, b}	Time (in s) for Version A - only numbers)	0–300
	TMT B ^{a, b}	Time (in s) for Version B - numbers and letters	0–500
	TMT B/A ^{a, b}	Calculated as $\frac{TMT\ B}{TMT\ A}$	0–
Verbal Fluency S-Words	Fluency S-Words	Number of correct answers	0–

Note. If the upper limit of the range is missing, the test score does not have an explicit upper limit. MMSE = Mini-Mental Status Examination; TMT = Trail Making Test.

^aThe outcome is a continuous variable. All other scores are discrete variables.

^bGreater test scores indicate worse performance. For all other scores greater test scores equal better performance.

Table 4. Results of the analysis of variance for the main effects of sex and education

Score	Sex			Education		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Fluency Animals	3.080	.081	.015	4.820	.029	.024
Boston Naming	0.001	.979	.000	0.544	.462	.003
MMSE	1.406	.237	.007	19.710	<.001	.091
Wordlist Learning						
List 1	3.137	.078	.016	0.002	.966	.000
List 2	14.679	<.001	.070	1.217	.271	.006
List 3	13.430	<.001	.064	2.463	.118	.012
Total	15.617	<.001	.074	1.325	.251	.007
Constructional Praxis						
Copying	0.329	.567	.002	12.833	<.001	.061
Wordlist Recall						
Recall	11.342	.001	.055	0.303	.583	.002
Savings	2.044	.154	.010	0.436	.510	.002
Intrusions	0.073	.788	.000	1.627	.204	.008
Wordlist Recognition	1.708	.193	.009	0.838	.361	.004
Constructional Praxis						
Recall	0.741	.390	.004	8.977	.003	.044
Savings	0.728	.394	.004	2.672	.104	.013
TMT A	3.329	.070	.017	3.279	.072	.016
TMT B	9.149	.003	.045	3.421	.066	.017
TMT B/A	2.277	.133	.011	0.039	.844	.000
Fluency S-Words	13.061	<.001	.062	1.309	.254	.007

Note. Test statistics for all Wordlist Learning scores, Wordlist Recall scores, TMT B, and TMT B/A are with (1,196) degrees of freedom. For all other scores, degrees of freedom were (1,197). The direction of the effect was the same for all significant effects. Females performed better than males and the high education group performed better than the less educated.

hearing performance were then included as predictors for each CERAD-NP score in order to test whether they were related to performance beyond the effects of the demographic variables. Results were only reported when a significant R^2 change was obtained from including sensory predictors. Data from one participant were excluded because no visual acuity test data were available.

RESULTS

The ANOVA (Table 4) revealed that the highly educated group performed better at Fluency Animals, MMSE, Constructional Praxis Copying and Recall. There were trends in the same direction for TMT A and TMT B. In addition, females performed better than males in the Wordlist Learning task (List 2, 3, and Total), Wordlist Recall, TMT B, and Fluency S-words. A trend level effect in the same direction was found for performance in Fluency Animals, Wordlist Learning List 1, and TMT A. The results for females and males did not differ significantly for any other scores. Due to these results norms were reported stratified according to gender and level of education.

An overview of the performance in each test score and the distribution of the data in the normative sample can be obtained from Table 5. Data from one person were missing for TMT B because this person did not want to complete it. Data from another person were retrospectively excluded from the analysis for the tasks Wordlist Learning, Recall, and Recognition because a Wordlist Total score of 2 (the next worst score in the overall sample was 9) indicated a lack of motivation during the learning trials. The normative data (subdivided according to sex and education) are presented in detail for each CERAD-NP score in a separate table in the supplement. The data in each table are presented from worse to better scores for easier interpretation. The discussion includes an example of how to use these reference tables.

Influence of Hearing and Vision

For the majority of CERAD-NP scores (16 out of 18), performance was not related to either hearing or visual acuity. However, visual acuity predicted performance in TMT A and B (Table 6). In all cases, worse visual acuity (indicated by larger logMAR) was related to worse task performance (more time for TMT A and B). Hearing performance predicted only TMT B. Again, hearing loss (indicated by less correctly repeated words) was associated with deficits in task performance (more time for TMT B). The estimates for the regression coefficients are presented in Table 6 and can be used to derive practical implications. For example, an increase of 1 in the logMAR scale means the time needed for the TMT A increases by approximately 32 s and for TMT B by approximately 86 s.

DISCUSSION

This study aimed to present normative data for all scores derived from the extended CERAD-NP for a sample of nondemented adults between 80 and 84 years of age. The normative data were presented as PR for discrete scores and as percentiles for continuous test scores and can be used as a reference point for performance of the old-old taking into consideration sex and educational level. Moreover, the effect of visual acuity and hearing on test performance was studied and indicated good robustness towards corrected sensory impairments. Only performance in the TMT was shown to suffer from lower visual acuity.

As shown in many previous studies, the demographic variables education and sex significantly influence CERAD-NP performance (e.g., Beeri et al., 2006; Kirsebom et al., 2019). Higher levels of education positively bias the performance (D. Y. Lee et al., 2004; Luck et al., 2018; Welsh et al., 1994). This was replicated in our sample and our highly educated group performed significantly better than the less educated group in the Fluency Animals, MMSE, Constructional Praxis Copying and Recall tests. The report of sex differences in CERAD-NP performance is not quite as one-sided, but seems to be more in favor of women performing better than their male counterparts of the same age (Beeri et al., 2006; Luck et al., 2018; McCurry et al., 2001). Females in our sample also performed better than males in a number of scores encompassing a wide variety of cognitive functions (language skills, memory, and executive functions). Males did not score significantly better than women in any of the test scores. Taken together these findings support the use of education- and sex-specific norms in neuropsychological testing, which is already common practice.

The validity of the data set was examined by comparing it with other normative data sets. This is only possible to a limited extent as reports often differ with regard to the exact characteristics of the study sample and calculations of norm values (Woerner et al., 2017). Nevertheless, we used data from Luck et al. (2009) to evaluate our data because age and nationality of both studies matched. In general, good agreement was found between both study samples. Their categorization of educational level with three categories (high, medium, low) differed slightly from our dichotomous categorization (<12 years vs. ≥ 12 years). This dichotomous variable makes our data set comparable internationally (Beeri et al., 2006; Nasreddine et al., 2005; Welsh-Bohmer, Gearing, Saunders, Roses, & Mirra, 1997; Welsh et al., 1994). Only the high and low education groups of Luck et al. (2009) were used for comparison as they had the most overlap with our groups. Scores available for comparison were: Fluency Animals, Wordlist Total, Wordlist Recall, Wordlist Recognition, and Wordlist Savings. Only the lower end of the data distribution was compared as this is decisive for the detection of impairments. Table 7 shows the highest score that is considered at least one standard deviation below the mean. The values were slightly higher in our sample with

Table 5. Descriptive data of all CERAD-NP scores for each of the four norm groups

	Male										Female									
	>12 years education (N = 75)					≤12 years education (N = 24) ^a					>12 years education (N = 47) ^b					≤12 years education (N = 55)				
	M	SD	Range	Skew	Kurt	M	SD	Range	Skew	Kurt	M	SD	Range	Skew	Kurt	M	SD	Range	Skew	Kurt
Fluency Animals	20.2	4.4	11–31	0.2	-0.3	18.4	4.9	11–29	0.4	-0.5	21.3	5.5	8–33	-0.6	0.1	19.8	4.6	12–33	0.7	0.4
Boston Naming	14.1	1.4	8–15	-2.1	5.3	14.0	1.1	11–15	-1.2	1.3	14.2	1.0	10–15	-1.8	5.1	13.9	1.3	9–15	-1.5	2.6
MMSE	28.1	1.3	24–30	-0.5	0.3	27.3	1.8	23–30	-0.3	-0.2	28.6	1.5	23–30	-1.7	3.9	27.3	1.7	24–30	-0.5	-0.7
Wordlist Learning																				
List 1	4.3	1.4	1–8	0.2	0.1	4.4	1.0	3–7	0.6	0.9	4.7	1.2	2–8	0.4	0.5	4.7	1.4	2–7	-0.2	-0.7
List 2	5.9	1.3	1–8	0.4	-0.2	5.6	0.9	4–7	0.3	-0.9	6.6	1.4	2–9	-0.7	1.1	6.4	1.5	3–10	-0.2	0.1
List 3	6.9	1.5	3–10	-0.2	0.0	6.7	1.2	5–9	0.3	-0.6	7.9	1.2	5–10	-0.1	-0.2	7.4	1.8	3–10	-0.4	-0.5
Total	17.1	3.4	9–25	-0.1	-0.3	16.6	2.5	13–23	0.6	0.5	19.2	2.5	14–25	-0.1	-0.3	18.5	3.9	9–26	-0.3	-0.6
Constructional Praxis																				
Copying	10.9	0.4	9–11	-3.3	10.7	10.2	1.1	7–11	-1.6	2.7	10.6	0.9	7–11	-2.7	6.8	10.3	1.3	5–11	-2.3	5.1
Wordlist Recall																				
Recall	5.1	1.8	0–9	-0.2	0.2	5.17	1.8	2–9	0.0	0.1	6.3	2.1	1–10	-0.5	0.1	6.0	1.9	0–10	-0.3	0.3
Savings	74.1	23.6	0–133	-0.3	0.6	77.4	20.8	29–100	-1.3	1.2	80.3	24.3	13–114	-0.8	0.6	81.9	25.8	0–150	0.0	1.5
Intrusions	0.5	1.2	0–6	3.0	9.1	0.8	1.4	0–4	1.4	1.1	0.6	1.3	0–5	2.4	5.5	0.9	1.6	0–9	3.3	13.6
Wordlist Recognition	93.1	6.5	70–100	-1.2	1.7	92.2	10.7	50–100	-3.1	11.2	94.9	7.4	65–100	-2.1	5.4	93.5	8.6	55–100	-2.2	6.6
Constructional Praxis																				
Recall	9.2	2.1	2–11	-1.3	1.8	8.0	2.2	4–11	-0.2	-0.7	8.8	2.4	2–11	-1.1	1.0	7.8	2.4	2–11	-0.5	-0.3
Savings	84.2	19.3	18–110	-1.3	1.9	79.6	24.7	40–157	1.0	3.1	82.1	20.8	22–113	-1.0	0.6	76.1	22.4	20–122	-0.4	-0.3
TMT A	57.0	20.7	30–149	1.7	4.4	70.7	29.9	30–180	2.3	7.5	58.2	19.2	27–106	0.6	0.1	57.0	23.4	28–155	1.9	5.3
TMT B	136.2	59.2	60–300	1.2	0.7	164.9	65.3	75–300	0.2	-1.0	121.0	55.5	55–300	2.2	5.3	125.6	56.9	20–300	1.5	2.5
TMT B/A	2.6	1.2	1.0–6.1	1.4	1.6	2.4	0.9	1.3–4.8	1.1	1.1	2.2	0.8	1.0–5.0	1.5	2.9	2.4	0.9	0.2–4.8	0.5	0.2
Fluency S-Words	13.2	3.9	3–24	0.0	0.4	12.1	3.4	3–18	-0.4	1.4	15.2	4.7	6–27	0.4	-0.1	14.8	4.5	4–24	0.0	-0.4

Note. Only data printed in bold are normally distributed according to Shapiro-Wilk test with α -level 0.05. Skew = Skewness, Kurt = Kurtosis, MMSE = Mini-Mental Status Examination, TMT = Trail Making Test.

^aOne person was rejected as an outlier for all Wordlist tasks and scores were calculated from N = 23.

^bOne person did not complete TMT B. The scores TMT B and TMT B/A were calculated from N = 46.

Table 6. Hierarchical multiple linear regression analysis to show the influence of visual acuity and hearing on cognitive performance

Variable	Model 1 – TMT A				Model 2 – TMT A			
	<i>B</i>	<i>SE</i>	β	<i>p</i>	<i>B</i>	<i>SE</i>	β	<i>p</i>
Age	1.8	1.2	.11	.14	0.8	0.5	.05	.52
Sex ^a	−3.5	3.5	−.08	.32	−1.0	3.4	−.02	.77
Education ^b	−0.4	0.6	−.06	.46	0.1	0.5	.01	.90
Visual Acuity					32.3	7.5	.31	<.001
Hearing					−0.4	0.3	−.10	.18
Corrected R ²				.02				.12

Variable	Model 1 – TMT B				Model 2 – TMT B			
	<i>B</i>	<i>SE</i>	β	<i>p</i>	<i>B</i>	<i>SE</i>	β	<i>p</i>
Age	1.4	3.2	.03	.68	−1.8	3.1	−.04	.54
Sex ^a	−27.1	9.1	−.23	.003	−21.2	8.7	−.18	.02
Education ^b	−1.6	1.4	−.16	.05	−1.6	1.4	−.08	.26
Visual Acuity					86.2	19.0	.31	<.001
Hearing					−1.9	0.8	−.17	.02
Corrected R ²				.05				.18

Note. Only models, where inclusion of hearing and visual acuity significantly improved R² are included in this table. TMT = Trail Making Test.

^a0 = male, 1 = female.

^bnumber of years.

Table 7. Highest score at least 1 SD below the norm for each group (comparison to Luck et al. (2009))

Score	Sex	Education	Luck et al. (2009)	SENDA	Difference
Fluency Animals	M	High	14	15	1
		Low	13	13	0
	F	High	14	16	2
		Low	12	14	2
Wordlist Total	M	High	13	14	1
		Low	11	13	2
	F	High	14	16	2
		Low	13	14	1
Wordlist Recall	M	High	2	3	1
		Low	1	3	2
	F	High	3	4	1
		Low	2	4	2
Wordlist Savings	M	High	40%	42%	2%
		Low	29%	33%	4%
	F	High	35%	49%	14%
		Low	29%	50%	21%
Wordlist Recognition	M	High	80%	85%	5% $\hat{=}$ 1 word
		Low	75%	85%	10% $\hat{=}$ 2 words
	F	High	85%	90%	5% $\hat{=}$ 1 word
		Low	80%	85%	5% $\hat{=}$ 1 word

Note. M = male, F = female, High = educational level > 12 years of education, Low = educational level \leq 12 years of education.

a 0–2 absolute point difference. One reason for this discrepancy could be the performance advantage of university-based samples compared to community-based samples, as this advantage remains even after controlling for educational level (Andel et al., 2003). University-based samples include volunteers who sign up for longitudinal studies with multiple visits at the university (comparable to the SENDA study). In contrast, community-based samples are recruited directly

in the community at senior centers (Andel et al., 2003) or from primary care facilities (Luck et al., 2009). Furthermore, these differences might be caused by the quality of education, an influencing factor on late-life cognition and health (Barba et al., 2021; Carvalho et al., 2015). Although samples with the same educational level were compared, it is unclear whether the quality of education was also comparable. A second reason for this discrepancy might be the

restricted age range (80–84 years) in our sample compared to the open-ended (> 80) category used by Luck et al. (2009), which resulted in a maximum participant age of 98 years. Including only a restricted age range decreases the risk of false-positive results for participants at the upper end of the age range. It has also been shown that even small differences in age can lead to significant differences in average performance (Beeri et al., 2006; Miller et al., 2015). For example, a group of 80–84 year-olds performed better than a group of 85–89 year-olds, which again differed from a group of 90–95 year-olds (Miller et al., 2015).

Comparison with an adjacent younger age group (75–79 years) from a recent population-based study in Germany (Luck et al., 2018) further supported the validity of our data set. When comparing means and 1 SD cut-offs, the younger group performed better than the present sample (80–84 years) across most scores, which confirms the negative relationship between age and cognitive performance. For Boston Naming, MMSE, Constructional Praxis Copying, Constructional Praxis Recall, and Constructional Praxis Saving, these differences only had a range of one point. In all other scores, the differences were even more pronounced. For example, the performance of a highly educated woman would be considered one standard deviation below the mean in the Fluency Animals score if they named 19 or fewer animals according to Luck et al. (2018). In comparison, the age-appropriate data presented here suggest the same cut-off is at 16 points. This further illustrates the increased risk of false positives when a younger reference group is used, even when the age differences (in this case 75–79 vs. 80–84 years) are relatively small (Beeri et al., 2006; Luck et al., 2018; Miller et al., 2015). An exception was the Fluency S-Words score where across all educational levels the older participants achieved slightly higher scores. It has been suggested that verbal fluency might be less affected by age because it reflects crystallized abilities like vocabulary and knowledge (Beeri et al., 2006).

Last, the data were compared to norms of nondemented volunteers of the directly following age range (85–89 years) from a US study (Beeri et al., 2006). It was expected that our sample would perform similarly or better because of their younger age. Comparing the highest score that was considered at or below the 10th percentile showed that for Boston Naming, MMSE, Wordlist Learning List 1, List 2, List 3, Total and Recall the values were either the same or within one point. For Fluency Animals, Constructional Praxis Copying, TMT A, and TMT B the differences were much more pronounced and always showed worse performance in the older age group. Somewhat surprising is the big drop-off in the Constructional Praxis Copying task (i.e., 10th percentile for highly educated males was ≤ 5 (Beeri et al., 2006) vs. ≤ 8 in the SENDA sample). Considering that the testing of the older sample was carried out at participant's homes instead of during lab visits, it is possible that more participants with movement restrictions (including fine motor impairments) were included.

Taken together, the comparisons presented above demonstrate that results from the SENDA study fit well into the previously published data. In addition, these findings are a valuable addition to the existing literature because they included all scores, instead of a small selection, and we provide clinically relevant percentiles (related to 1, 1.5, and 2.0 SD).

The following example illustrates how the normative tables provided in the supplement can be used in practice. For the sample case (woman, 83 years old, 14 years of education) the following performances were recorded: Wordlist Learning Total–17, Wordlist Savings–35%. The Wordlist Total score is discrete and, hence, the number of points (17) must be looked up in the first column. In the same row, we find a PR of 21.3% in the column “Female >12 years education,” which means that 21.3% of the reference sample scored the same or fewer points. This is also equal to a z-score of -0.8 , which indicates that the performance was below average but did not reach the -1.5 SD cut-off usually used to determine mild cognitive impairments. In contrast, the Wordlist Savings score is continuous and must be compared to the numbers given in the column “Female > 12 years education.” Looking for the closest number above the score reached (35%), we find 40%, which is equal to the 6.68th percentile and $z = -1.50$. From this, we know that less than 6.68% of the reference sample performed worse than the sample case and that this test performance is more than 1.5 SD below the reference average indicating impairment in recall performance.

A further argument for providing this new data set for individuals 80–84 years of age is that neuropsychological reference data are ideally kept up-to-date to deal with cohort effects and socio-environmental changes that can alter typical test performance (Dickinson & Hiscock, 2011). The Flynn-Effect describes the phenomenon of generational gains in intelligence testing, which means that later-born cohorts typically have higher levels of fluid intelligence (Flynn, 1987; Skirbekk, Stonawski, Bonsang, & Staudinger, 2013). Similarly, it has been shown that the performance of OA in processing speed, language, executive function, and verbal memory tasks has improved across birth cohorts and that this trend could be ongoing in the future (Dodge et al., 2017; Skirbekk et al., 2013). Therefore, it is important to publish data shortly after data collection and to also include information about birth cohorts, as was done here. Using outdated references could potentially lead to missing cases of cognitive impairment or limit us to only being able to detect them later in the transition to disease.

Some limitations of this sample must be considered. Establishing the dementia-free status was based on self-report (no diagnosed dementia) and neuropsychological screening, but did not include a full clinical assessment. Therefore, the inclusion of as yet undetected cases of dementia cannot be completely ruled out. However, the number of such cases should be minimal because, in addition to the participant's self-report of clinical diagnoses, performance in the MoCA

was used to exclude such cases. Another potential limitation may arise because the birth cohort included in the sample (born between 1933 and 1939) grew up in Germany in the aftermath of World War II (1939–1945). This has been shown to have long-lasting effects on health and lifestyle into old age (Conzo & Salustri, 2019; Havari & Peracchi, 2017). In addition, all participants were current residents of Chemnitz and its surroundings and the vast majority of them lived in eastern Germany all their life. From this follows a very distinct difference in the socialization conditions during their working adulthood in the GDR (German Democratic Republic) compared to people who lived in the FRD (Federal Republic of Germany). This may result in the sample not being representative for the whole German population of this age group. Comparisons between East and West German OA have shown that East German women perform better in memory and fluid intelligence tests compared to their West German counterparts (Rupprecht, 2000). It is assumed that this effect is caused by the higher rate of employment for women in the GDR (Rupprecht, 2000). Beyond this, a bias during recruitment cannot be excluded, which probably favored more educated and healthier adults. The final limitation relates to sample size, sample sizes of $N = 50$ – 75 are considered a sufficient compromise between costs for data acquisition and generalizability for neuropsychological tests norms (Bridges & Holler, 2007). The group of men with less than 12 years of education is relatively small ($n = 24$) compared to the other groups, which all meet this recommendation. Other studies also reported problems finding enough male participants with a low educational level (Beerli et al., 2006; Welsh et al., 1994). As there are no current and complete reference values for this age group, this sample must still be considered a valuable expansion of the existing data.

Influence of Hearing and Vision

As a secondary outcome, we were also interested in whether CERAD-NP performance might be related to hearing ability and/or visual acuity, even in a sample of nondemented participants suffering from an age-typical decline in vision and/or hearing. The results indicate that most CERAD-NP subtests are robust regarding the age-related sensory loss found in an old-old age group. This reinforces the good practical application of the test battery. It should be considered that all participants were asked to use vision and hearing aids during testing. Hence, this does not mean that sensory performance per se is irrelevant for test performance. Rather it suggests that as long as no pathological visual or hearing impairments are present, the tasks can be conducted adequately.

Nevertheless, visual acuity predicted the TMT A, and TMT B scores in our sample. The TMT is a visual search paradigm, where 25 letters (TMT A) or 13 letters and 12 numbers (TMT B) are distributed over a sheet of paper and must be connected in the correct order. Hence, the negative effect of visual acuity loss (even when corrected) on performance time is not surprising. This is also in accordance

with findings that patients with glaucoma performed worse in the TMT B (S. S. Lee, Wood, & Black, 2020). Therefore, the time needed to perform TMT A and B must be interpreted with caution. Fortunately, the third score TMT B/A, which is the quotient of both times, showed no relationship to visual acuity. As the visual search demands of both conditions are similar, the slowing in both due to visual impairments seems to cancel out. In addition, the TMT B/A was found to be a purer measure of executive functions (Arbuthnott & Frank, 2010) and to be less susceptible to effects of demographics (Christidi, Kararizou, Triantafyllou, Anagnostouli, & Zalonis, 2015). In summary, this supports the utilization of the TMT B/A score.

Only a single score (TMT B) was associated with hearing ability, although less so than with visual acuity. Worse hearing performance was related to longer times needed for the TMT B. This relationship seems counterintuitive, as there is no hearing involved in solving this task. However, the TMT B is known to be one of the more difficult tasks involving a high cognitive load. As a result, a high number of participants with dementia are unable to complete it (Schmid et al., 2014). Therefore, the relationship with hearing ability could be caused by fatigue, which, according to the effortfulness hypothesis, would be more severe in individuals with worse hearing as they would expend a lot more effort across the whole testing session understanding the oral instructions. The common cause hypothesis may also explain the relationship between hearing ability and task performance in tasks with no auditory stimuli (TMT B) indicating that the sensory and cognitive system were affected by the same neuropathological processes. In summary, lower sensory performance seems to be a concern for TMT performance, but the alternative scoring option (TMT B/A) can be used instead. It is noteworthy that all participants used glasses as needed and were, in general, considered to have normal to moderate-low corrected vision. The negative effects were present even though participants did not report any problems with the visual stimuli.

To conclude, this data set of nondemented individuals with an age between 80 and 84 years presents reference data for the application of the CERAD-NP in this age group in Germany and any population similar to the sample described here. The normative tables presented include all information required to easily evaluate test scores in comparison to the typical performance of this age group, while also taking into consideration sex and educational level. This will help improve the diagnostic process of dementia in old-old age because individuals that should be remitted for further diagnostics can be identified. In the future, these references will need to be supplemented by additional normative data sets that include individuals 85 years and older in order to cover the entire age spectrum for neuropsychological testing.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <https://doi.org/10.1017/S1355617721001284>

FINANCIAL SUPPORT

This work was funded by the European Social Fund for Germany and the Sächsische Aufbaubank-Förderbank (SAB) of the Free State of Saxony (Project-Number: 100310502, address: Pirnaische Straße 9, 01069 Dresden, Germany).

CONFLICT OF INTEREST

The authors report no potential conflicts of interest.

REFERENCES

- Aebi, C. (2002). Validierung der neuropsychologischen Testbatterie CERAD-NP: Eine Multi-Center Studie [Validation of the neuropsychological test battery CERAD-NP: A multi-center study]. Doctoral Dissertation, University of Basel.
- Andel, R., McCleary, C.A., Murdock, G.A., Fiske, A., Wilcox, R.R., & Gatz, M. (2003). Performance on the CERAD word list memory task: A comparison of university-based and community-based groups. *International Journal of Geriatric Psychiatry*, *18*(8), 733–739. doi: [10.1002/gps.913](https://doi.org/10.1002/gps.913).
- Arbuthnott, K. & Frank, J. (2010). Trail Making Test, part B as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 518–528. doi: [10.1076/1380-3395\(200008\)22:4;1-0:ft518](https://doi.org/10.1076/1380-3395(200008)22:4;1-0:ft518).
- Bach, M. (1996). The Freiburg Visual Acuity test—Automatic measurement of visual acuity. *Optometry and Vision Science*, *73*(1), 49–53.
- Barba, C., Garcia, A., Clay, O.J., Wadley, V.G., Andel, R., Davila, A.L., & Crowe, M. (2021). Quality of education and late-life cognitive function in a population-based sample from Puerto Rico. *Innovation in Aging*, *5*(2). doi: [10.1093/geroni/igab016](https://doi.org/10.1093/geroni/igab016).
- Beeri, M.S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H., & Silverman, J.M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, *67*(6), 1006–1010. doi: [10.1212/01.wnl.0000237548.15734.cd](https://doi.org/10.1212/01.wnl.0000237548.15734.cd).
- Boyd, D. & Bee, H. (2006). *Lifespan Development* (4th ed.). London: Pearson.
- Breton, A., Casey, D., & Arnaoutoglou, N.A. (2019). Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *International Journal of Geriatric Psychiatry*, *34*(2), 233–242. doi: [10.1002/gps.5016](https://doi.org/10.1002/gps.5016).
- Bridges, A.J. & Holler, K.A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology*, *13*(6), 528–538. doi: [10.1080/09297040701233875](https://doi.org/10.1080/09297040701233875).
- Carvalho, J.O., Tommet, D., Crane, P.K., Thomas, M.L., Claxton, A., Habeck, C., ... Romero, H.R. (2015). Deconstructing racial differences: The effects of quality of education and cerebrovascular risk factors. *Journal of Gerontology: Series B*, *70*(4), 545–556. doi: [10.1093/geronb/gbu086](https://doi.org/10.1093/geronb/gbu086).
- Charlson, M.E., Pompei, P., Ales, K.L., & MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383. doi: [10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8).
- Christidi, F., Kararizou, E., Triantafyllou, N., Anagnostouli, M., & Zalonis, I. (2015). Derived Trail Making Test indices: Demographics and cognitive background variables across the adult life span. *Aging, Neuropsychology, and Cognition*, *22*(6), 667–678. doi: [10.1080/13825585.2015.1027650](https://doi.org/10.1080/13825585.2015.1027650).
- Colenbrander, A. (2002). Visual standards: aspects and ranges of vision loss with emphasis on population surveys. *Report for the International Council of Ophthalmology*, 1–33.
- Conzo, P. & Salustri, F. (2019). A war is forever: The long-run effects of early exposure to World War II on trust. *European Economic Review*, *120*. doi: [10.1016/j.euroecorev.2019.103313](https://doi.org/10.1016/j.euroecorev.2019.103313).
- Crawford, J., Garthwaite, P., & Slick, D. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, *23*, 1173–1195. doi: [10.1080/13854040902795018](https://doi.org/10.1080/13854040902795018).
- Deutsche Alzheimer Gesellschaft e.V. (2018). Informationsblatt 1 Die Häufigkeit von Demenzerkrankungen [Factsheet 1 The prevalence of dementia]. Retrieved from https://www.deutsche-alzheimer.de/fileadmin/alz/pdf/factsheets/infoblatt1_haeufigkeit_demenzerkrankungen_dalzg.pdf
- Dickinson, M.D. & Hiscock, M. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology*, *18*(2), 136–142. doi: [10.1080/09084282.2010.547785](https://doi.org/10.1080/09084282.2010.547785).
- Diener, E., Emmons, R.A., Larsen, R.J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. doi: [10.1207/s15327752jpa4901_13](https://doi.org/10.1207/s15327752jpa4901_13).
- Dodge, H.H., Zhu, J., Hughes, T.F., Snitz, B.E., Chang, C.H., Jacobsen, E.P., & Ganguli, M. (2017). Cohort effects in verbal memory function and practice effects: A population-based study. *International Psychogeriatrics*, *29*(1), 137–148. doi: [10.1017/S1041610216001551](https://doi.org/10.1017/S1041610216001551).
- Duke Aging Center. (n.d.). CERAD – Consortium to Establish a Registry for Alzheimer’s Disease – Translations. Retrieved from <https://sites.duke.edu/centerforaging/cerad/translations/>
- Dupuis, K., Pichora-Fuller, M.K., Chasteen, A.L., Marchuk, V., Singh, G., & Smith, S.L. (2015). Effects of hearing and vision impairments on the Montreal Cognitive Assessment. *Aging, Neuropsychology, and Cognition*, *22*(4), 413–437. doi: [10.1080/13825585.2014.968084](https://doi.org/10.1080/13825585.2014.968084).
- Fillenbaum, G.G., van Belle, G., Morris, J.C., Mohs, R.C., Mirra, S.S., Davis, P.C., ... Welsh-Bohmer, K.A. (2008). Consortium to Establish a Registry for Alzheimer’s Disease (CERAD): The first twenty years. *Alzheimer’s & Dementia*, *4*(2), 96–109.
- Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171.
- Folstein, M.F., Folstein, S.E., & 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*(3), 189–198. doi: [10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6).
- Gauggel, S. & Birkner, B. (1999). Validität und Reliabilität einer deutschen Version der Geriatrischen Depressionsskala (GDS) [Validity and reliability of a German version of the Geriatric Depression Scale (GDS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *28*(1), 18–27. doi: [10.1026/0084-5345.28.1.18](https://doi.org/10.1026/0084-5345.28.1.18).
- Hahlbrock, K.-H. (1953). Über Sprachaudiometrie und neue Wörtertteste [Speech audiometry and new word-tests]. *Archiv für Ohren-, Nasen- und Kehlkopfheilkunde*, *162*(5), 394–431.

- Havari, E. & Peracchi, F. (2017). Growing up in wartime: Evidence from the era of two world wars. *Economics and Human Biology*, 25, 9–32. doi: [10.1016/j.ehb.2016.09.002](https://doi.org/10.1016/j.ehb.2016.09.002).
- Hesse, G., Eichhorn, S., & Laubert, A. (2014). Hörfähigkeit und Schwerhörigkeit alter Menschen [Hearing function and hearing loss in the elderly]. *HNO*, 62(9), 630–639. doi: [10.1007/s00106-014-2903-8](https://doi.org/10.1007/s00106-014-2903-8).
- Kirsebom, B.-E., Espenes, R., Hessen, E., Waterloo, K., Johnsen, S.H., Gundersen, E., ... Fladby, T. (2019). Demographically adjusted CERAD wordlist test norms in a Norwegian sample from 40 to 80 years. *The Clinical Neuropsychologist*, 33(sup1), 27–39.
- Lee, D.Y., Lee, K.U., Lee, J.H., Kim, K.W., Jhoo, J.H., Kim, S.Y., ... Woo, J.I. (2004). A normative study of the CERAD neuropsychological assessment battery in the Korean elderly. *Journal of the International Neuropsychological Society*, 10(1), 72.
- Lee, S.S., Wood, J.M., & Black, A.A. (2020). Impact of glaucoma on executive function and visual search. *Ophthalmic and Physiological Optics*, 40(3), 333–342. doi: [10.1111/opo.12679](https://doi.org/10.1111/opo.12679).
- Li, K.Z.H. & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience & Biobehavioral Reviews*, 26(7), 777–783.
- Lienert, G.A. & Raatz, U. (1998). Die Testeichung [Test standardization]. In *Testaufbau und Testanalyse [Test construction and analysis]* (6th ed., pp. 272–296). Weinheim: Psychologie Verlags Union.
- Lim, M.Y.L. & Loo, J.H.Y. (2018). Screening an elderly hearing impaired population for mild cognitive impairment using Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA). *International Journal of Geriatric Psychiatry*, 33(7), 972–979. doi: [10.1002/gps.4880](https://doi.org/10.1002/gps.4880).
- Liu, K.P., Kuo, M.C., Tang, K.C., Chau, A.W., Ho, I.H., Kwok, M.P., ... Chu, L.W. (2011). Effects of age, education and gender in the Consortium to Establish a Registry for the Alzheimer's Disease (CERAD)-Neuropsychological Assessment Battery for Cantonese-speaking Chinese elders. *International Psychogeriatrics*, 23(10), 1575–1581. doi: [10.1017/S1041610211001153](https://doi.org/10.1017/S1041610211001153).
- Luck, T., Pabst, A., Rodriguez, F.S., Schroeter, M.L., Witte, V., Hinz, A., ... Thiery, J. (2018). Age-, sex-, and education-specific norms for an extended CERAD Neuropsychological Assessment Battery—Results from the population-based LIFE-Adult-Study. *Neuropsychology*, 32(4), 461.
- Luck, T., Riedel-Heller, S.G., Wiese, B., Stein, J., Weyerer, S., Werle, J., ... für die AgeCoDe Study, G. (2009). CERAD-NP-Testbatterie: Alters-, geschlechts- und bildungsspezifische Normen ausgewählter Subtests [CERAD-NP battery: Age-, gender- and education-specific reference values for selected subtests]. *Zeitschrift für Gerontologie und Geriatrie*, 42(5), 372–384. doi: [10.1007/s00391-009-0031-y](https://doi.org/10.1007/s00391-009-0031-y).
- McCoy, S.L., Tun, P.A., Cox, L.C., Colangelo, M., Stewart, R.A., & Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology*, 58(1), 22–33. doi: [10.1080/02724980443000151](https://doi.org/10.1080/02724980443000151).
- McCurry, S.M., Gibbons, L.E., Uomoto, J.M., Thompson, M.L., Graves, A.B., Edland, S.D., ... Larson, E.B. (2001). Neuropsychological test performance in a cognitively intact sample of older Japanese American adults. *Archives of Clinical Neuropsychology*, 16(5), 447–459. doi: [10.1093/arclin/16.5.447](https://doi.org/10.1093/arclin/16.5.447).
- Memory Clinic Basel. (2005). *Manual zum Auswertungsprogramm CERAD-Plus 1.0 [Scoring manual for the CERAD-Plus 1.0]*. Basel: Memory Clinic.
- Miller, I.N., Himali, J.J., Beiser, A.S., Murabito, J.M., Seshadri, S., Wolf, P.A., & Au, R. (2015). Normative data for the cognitively intact oldest-old: The Framingham Heart Study. *Experimental Aging Research*, 41(4), 386–409. doi: [10.1080/0361073X.2015.1053755](https://doi.org/10.1080/0361073X.2015.1053755).
- Morris, J.C., Edland, S., Clark, C., Galasko, D., Koss, E., Mohs, R., ... Heyman, A. (1993). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD): Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology*, 43(12), 2457.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159–1165.
- Müller, K., Fröhlich, S., Germano, A.M.C., Kondragunta, J., Agoitia Hurtado, M.F.D.C., Rudisch, J., ... Voelcker-Rehage, C. (2020). Sensor-based systems for early detection of dementia (SENDA): A study protocol for a prospective cohort sequential study. *BMC Neurology*, 20(1), 84. doi: [10.1186/s12883-020-01666-8](https://doi.org/10.1186/s12883-020-01666-8).
- Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699. doi: [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x).
- Nowak, J.Z. (2006). Age-related macular degeneration (AMD): Pathogenesis and therapy. *Pharmacological Reports*, 58(3), 353.
- O'Bryant, S.E., Edwards, M., Johnson, L., Hall, J., Gamboa, A., & O'jile, J. (2018). Texas Mexican American adult normative studies: Normative data for commonly used clinical neuropsychological measures for English- and Spanish-speakers. *Developmental Neuropsychology*, 43(1), 1–26. doi: [10.1080/87565641.2017.1401628](https://doi.org/10.1080/87565641.2017.1401628).
- Oberg, M., Marcusson, J., Nagga, K., & Wressle, E. (2012). Hearing difficulties, uptake, and outcomes of hearing aids in people 85 years of age. *International Journal of Audiology*, 51(2), 108–115. doi: [10.3109/14992027.2011.622301](https://doi.org/10.3109/14992027.2011.622301).
- Paajanen, T., Hanninen, T., Tunnard, C., Mecocci, P., Sobow, T., Tsolaki, M., ... AddNeuroMed, C. (2010). CERAD neuropsychological battery total score in multinational mild cognitive impairment and control populations: the AddNeuroMed study. *Journal of Alzheimer's Disease*, 22(4), 1089–1097. doi: [10.3233/JAD-2010-100459](https://doi.org/10.3233/JAD-2010-100459).
- Reitmeir, P., Linkohr, B., Heier, M., Molnos, S., Strobl, R., Schulz, H., ... Graw, J. (2017). Common eye diseases in older adults of southern Germany: results from the KORA-Age study. *Age and Ageing*, 46(3), 481–486. doi: [10.1093/ageing/afw234](https://doi.org/10.1093/ageing/afw234).
- Rupperecht, R. (2000). Kognitive Leistungsfähigkeit im Kohortenvergleich [Cohort comparisons of cognitive performance]. In P. Martin, K.U. Ettrich, U. Lehr, D. Roether, M. Martin, & A. Fischer-Cyrules (Eds.), *Aspekte der Entwicklung im mittleren und höheren Lebensalter: Ergebnisse der Interdisziplinären Längsschnittstudie des Erwachsenenalters (ILSE)* (pp. 68–82). Heidelberg: Steinkopff.

- Schmid, N.S., Ehrensperger, M.M., Berres, M., Beck, I.R., & Monsch, A.U. (2014). The extension of the German CERAD neuropsychological assessment battery with tests assessing subcortical, executive and frontal functions improves accuracy in dementia diagnosis. *Dementia and Geriatric Cognitive Disorders Extra*, 4(2), 322–334. doi: [10.1159/000357774](https://doi.org/10.1159/000357774).
- Skirbekk, V., Stonawski, M., Bonsang, E., & Staudinger, U.M. (2013). The Flynn effect and population aging. *Intelligence*, 41(3), 169–177. doi: [10.1016/j.intell.2013.02.001](https://doi.org/10.1016/j.intell.2013.02.001).
- Statistisches Bundesamt (Destatis). (2019). *Bevölkerung Deutschlands bis 2060: Ergebnisse der 14. koordinierten-Bevölkerungsvorausberechnung – Hauptvarianten 1 bis 9 [Population of Germany until 2060: Variants and model calculations of the 14th coordinated population projection]*. Wiesbaden, Germany: Federal Statistical Office (Destatis).
- Tsai, J.C. (2009). A comprehensive perspective on patient adherence to topical glaucoma therapy. *Ophthalmology*, 116(11 Suppl), S30–S36. doi: [10.1016/j.ophtha.2009.06.024](https://doi.org/10.1016/j.ophtha.2009.06.024).
- Uchida, Y., Sugiura, S., Nishita, Y., Saji, N., Sone, M., & Ueda, H. (2019). Age-related hearing loss and cognitive decline – The potential mechanisms linking the two. *Auris Nasus Larynx*, 46(1), 1–9. doi: [10.1016/j.anl.2018.08.010](https://doi.org/10.1016/j.anl.2018.08.010).
- Welsh, K.A., Butters, N., Mohs, R.C., Beekly, D., Edland, S., Fillenbaum, G.G., & Heyman, A. (1994). The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology*, 44(4), 609.
- Welsh-Bohmer, K.A., Gearing, M., Saunders, A.M., Roses, A.D., & Mirra, S. (1997). Apolipoprotein E genotypes in a neuropathological series from the Consortium to Establish a Registry for Alzheimer’s Disease. *Annals of Neurology*, 42(3), 319–325. doi: [10.1002/ana.410420308](https://doi.org/10.1002/ana.410420308).
- Woerner, W., Müller, C., & Hasselhorn, M. (2017). Bedeutung und Berechnung der Prozentränge und T-Werte beim Erstellen von Testnormen: Anmerkungen und Empfehlungen [Meaning and calculation of percentile ranks and T-values when generating test standards: Remarks and recommendations. Gifts and talents]. *Begabungen und Talente. Jahrbuch der pädagogisch-psychologischen Diagnostik, Tests & Trends*, 15, 245–263.
- Wood, J., Chaparro, A., Anstey, K., Lacherez, P., Chidgey, A., Eisemann, J., . . . La, P. (2010). Simulated visual impairment leads to cognitive slowing in older adults. *Optometry and Vision Science*, 87(12), 1037–1043. doi: [10.1097/OPX.0b013e3181fe64d7](https://doi.org/10.1097/OPX.0b013e3181fe64d7).