universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Neuropsychological prediction of dementia using the Neuropsychological Test Battery Vienna – A retrospective study"

verfasst von / submitted by

Anna Catalina Garcia Rosas, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG 2002
Betreut von / Supervisor:	Univ. Prof. Dr. Claus Lamm

Die Masterarbeit wäre ohne die Unterstützung von einigen Personen nicht zum Abschluss gekommen. Dafür möchte ich mich an dieser Stelle ganz herzlich bedanken.

Als erstes möchte ich mich bei Assoc. Prof. Priv. Doz. Mag. Dr. Lehrner für die Bereitstellung des Themas und die spannende Zeit im AKH während der Datenerhebung bedanken. Ebenso möchte ich mich bei Univ. Prof. Dr. Lamm bedanken, der die offizielle Betreuung meiner Masterarbeit übernommen hat und mit seinem Fachwissen, meine Expertise im neuropsychologischen Forschungsbereich erweitert hat.

Während meiner Arbeit mit demenzkranken Patienten ist mir umso mehr bewusst geworden, wie wichtig die Angehörigen und das soziale Umfeld im Leben sind. Sie haben in den Bereichen geholfen, die die Betroffenen nicht mehr allein bewältigen konnten und ihnen Rückhalt gegeben, um die Lebensqualität aufrecht zu erhalten. Mit der Arbeit ist mir nochmal deutlich geworden, dass ich mich glücklich schätzen kann so einen Halt auch in meinem Leben zu haben. Ein ganz besonderer Dank geht dafür an meine Freunde, die mich vor allem in dieser stressigen Zeit begleitet haben. Durch eure liebevolle und lebensfrohe Art habt ihr mich mit Ratschlägen, Zuspruch, Zuhören, Aufmunterungen und so viel mehr unterstützt und damit meine Leben positiv beeinflusst. Ihr gebt mir das Gefühl immer für mich da zu sein und ich kann mich in jeder Situation auf euch verlassen. Anna-Maria, Becci, Dariusch, Dina, Felix, Kati, Lena, Luki, Maya, Resi, Vera - ich kann euch gar nicht genug danken.

Ebenso danke ich meiner Familie, die mich trotz der weiten räumlichen Distanz unterstützt haben. Danke an meine Geschwister Alina, Basir und René, bei denen ich weiß, dass ich zu jedem Zeitpunkt auf sie zählen kann. Mein größter Dank geht an meine Eltern, ohne die ich nicht an dem Punkt in meinem Leben stehen würde, wo ich mich jetzt befinde. Ihr steht immer hinter mir, egal welche Entscheidungen ich treffe. Ihr seid für mich da, wenn ich allein nicht mehr weiterkomme und bestärkt mich in jeder Hinsicht, um mir ein glückliches Leben zu ermöglichen. Danke, dass ihr mich meinen Weg gehen lässt und mir die nötige Unterstützung dafür gebt!

Table of Contents

Introduction

According to recent studies, the proportion of elderly people within the population has increased steadily in recent years and will continue to rise. This is due to declining birth rates and increased life expectancy fostered by medical progress and developments in the healthcare sector (Kaindl & Schipfer, 2017). However, despite advances in the medical field, the prevalence of diseases increases with age. Especially dementia is very common among older people (Berr, Wancata, & Ritchie, 2005; Reynish, Bickel, Fratiglioni, Kiejna, Prince, & Georges, 2009). About 115 000 to 130 000 people with a particular form of dementia currently live in Austria. This number is expected to double by 2050 due to the aging population (Höfler, Bengough, Winkler, & Griebler, 2015). There are different aetiologies of dementia disorders. The most typical form is Alzheimer's disease (AD), accounting for 60.0 – 80.0 % of dementias. Furthermore, there is vascular dementia (15.0 - 20.0 %) and the dementia with Lewy bodies (7.0 – 20.0 %). While other forms of dementia are rather rare (< 10.0 %), hybrids are more frequent (Scheichenberger, 2018). Dementia is characterized by acquired disorders of cognitive performance and emotional control of conscious patients that interfere with everyday activities (Lehrner, Bodner, Dal-Bianco, & Schmidt, 2011). The typical development of dementia involves three phases. The first phase contains the subjective cognitive decline (SCD), which transitions to mild cognitive impairment (MCI) and ends in dementia (Jessen et al., 2010; Jessen et al., 2014a).

Subjective cognitive decline

SCD is a deterioration of cognitive functions expressed by people who have no deficits in objective testing (Mitchell, Beaumont, Ferguson, Yadegarfar, & Stubbs, 2014). It is influenced by several factors like aging, personality traits, psychiatric conditions, neurologic and medical disorders, substance abuse and medication (Boone, 2009; Jessen et al., 2014b). The criteria for diagnosing SCD are

1. Self-experienced persistent decline in cognitive capacity in comparison with a previously normal status and unrelated to an acute event.
2. Normal age-, gender-, and education-adjusted performance on standardized cognitive tests, which are used to classify mild cognitive impairment or prodromal AD (Jessen et al., 2014b).

The exclusion criteria are mild cognitive impairment, prodromal AD or dementia. Therefore, the diagnose SCD can be explained by a psychiatric or neurologic disease apart from AD, medical disorder, medication or substance abuse (Jessen et al., 2014b).

Subjective memory complaints were associated with increased risk of development to dementia (Jessen et al., 2014a; Luck, Luppa, Matschinger, Jessen, Angermeyer, & Riedel-Heller, 2015; Reisberg, Shulman, Torossian, Leng, & Zhu, 2010) and there is evidence, that SCD occurs at the preclinical stage of AD (Glodzik-Sobanska et al., 2007).

Mild cognitive impairment

In the case of MCI, a subjectively perceived cognitive deficit occurs, which is reflected in impairments of poorer performance in the neuropsychological test (1.50 Standard deviation (*SD*) below age, gender and education specifics norms; Lehrner, Maly, Gleiss, Auff, & Dal-Bianco, 2008; Petersen, Stevens, Ganguli, Tangalos, Cummings, & DeKosky, 2001) but does not yet fulfil the criteria of dementia. Causes for this can include depressive mood, internal disorders and incipient dementia. Petersen (2004) developed criteria to diagnose the prodromal phase of AD, the MCI. The clinical criteria are:

1) Change in cognition observed by the patient, an informant, or qualified physician,
2) Change is untypical for age,
3) Impairment in one or more cognitive domains including memory, executive function, attention, language and visuo-constructional skills,
4) Decline in essential normal functional activities, like getting dressed by yourself and,
5) Absence of dementia (Petersen, 2004).

Because of the similar course of neuropathological changes, all AD patients undergo the typical phase progression from fully asymptomatic to subclinical phases until the onset of dementia. For MCI, an annual conversion rate to AD of 9.0 to 23.0 % is reported (Lehrner et al., 2005; Lehrner et al., 2016, Valencia & Lehrner, 2018).

Alzheimer's disease

AD is a slowly but steadily progressing and irreversible mental disorder and the most common type of dementia (Lehrner et al., 2011). The neurodegenerative disease of the human nervous system is characterized by two typical neuropathological changes. These neuropathological changes result in extracellular amyloid deposition of abnormally altered Aβ42-Protein and the formation of intraneuronal aggregates consisting of abnormal tau protein with subsequent neurofibrillary degeneration (Thal & Braak, 2005). These lead to the death of cerebral nerve cells with subsequent atrophy of the brain (Lehrner et al., 2011). Because of the loss of neurons and the shrinking of cerebral matter, the synaptic transmission, responsible for the transmission of information and the information processing, starts to fail with time (Kumar & Foster, 2007). As a result, the initial neuropathological manifestations also cause noticeable problems for affected patients. Persons with AD have shown preclinical deficits in several

cognitive domains (Wittchen & Hoyer, 2011). This includes memory disorders with symptoms like memorization problems or occasionally spatial orientation problems. The most steady and noticeable cognitive deficits in this disease are seen in functions related to the episodic memory (Baudic, Dalla Barba, Thibaudet, Smagghe, Remy, & Traykov, 2006; Bäckman, Small, & Fratiglioni, 2001). The episodic memory is part of the explicit memory, responsible for memorizing the personal autobiography and watershed events of the public life e.g. in politics or economy, which do not have an immediate effect on the personal life. Mainly the prefrontal cortex, parts of the parietal lobe, the limbic system and the cingulum control the episodic memory (Bear, Connors, & Paradiso, 2016). Due to the impairment of the episodic memory, AD patients can no longer or only fragmentarily remember situations of the last hours, days and weeks. Additionally, those who are affected forget notable events or appointments, sometimes whole situations and get lost in unknown surroundings (Lehrner et al., 2011). While the short-term memory starts to deteriorate, the ability to remember events from the past stays. This leads to a biographically reoriented life of the affected persons. Due to the memory disorder of the neurodegenerative disease, early constraints in the local and temporal orientation are typical. Later, the situational orientation and ultimately one's very own orientation are affected as well (Lehrner et al., 2011). In addition to memorization problems and memory disorders, semantic memory impairments are also noticeable in AD. These are expressed in word-finding disorders that affect, among other things, the name memory (Maseda, Lodeiro-Fernández, Lorenzo-López, Núñez-Naveira, Balo, & Millán-Calenti, 2014). In the early stages of AD, the speech content of those who are affected is reduced. Their sentences become shorter and the grammar turns faulty. Deficiencies are observed in the speech intelligibility, choice of words, writing and reading. The motor speech ability and the ability to repeat, by contrast, remain unchanged even in severe stages of dementia. Mental arithmetic often causes problems rather early (Lehrner et al., 2011). Generally, patients lose the most recently developed cognitive abilities at an early stage. They can no longer solve complex tasks; the focused attention for planning, abstract thinking and acting worsens continuously and the affected persons are easily distracted. This is often expressed for the first time in everyday activities such as driving or personal hygiene (Lehrner et al., 2011; Schiffczyk, Romero, Jonas, Lahmeyer, Müller, & Riepe, 2013). To diagnose AD, the following criteria must be met according to the Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5):

1) Memory impairment (impaired ability to learn new information or to recall previously learned information).
2) Aphasia and/or apraxia, agnosia or an impairment in executive functioning.

3) Deficits must include a significant impairment in social or occupational functioning and represent a decline from a previous level of performance e.g. not able to take a shower on your own.
4) The course of the disease is characterized by gradual onset and continuing cognitive decline.
5) The cognitive impairments in 1) and 2) exclude other psychiatric disorders or neurological explanations for the decline of function.
6) The cognitive impairments do not occur only during the course of a delirium (American Psychiatric Association, 2013).

In summary, a person who develops AD shows significant cognitive deficits, followed by behavioral, cognitive and social changes (Bennett, 1995; Jacobs, Sano, Dooneief, Marder, Bell, & Stern, 1995; Lehrner et al., 2011; Small, Herlitz, Fratiglioni, Almkvist, & Bäckman, 1997).

Dementia and cognitive deficits

Before the appearance of clinically significant cognitive change, the development of dementia can be predicted with some success based on changes in test scores resulting from specific neuropsychological tests, like the Neuropsychological Test Battery Vienna (NTBV). Several cognitive performances can be used as a predictor of dementia. The most salient predictor for incident dementia is the episodic memory. In particular, deficits of the verbal and visuospatial memory, are both one of the earliest signs which have been reported as being related to increased risk of developing dementia (Drummond et al., 2015; Elias, Beiser, Wolf, Au, White, & D'Agostino 2000; Grober, Lipton, Hall, & Crsytal, 2000; Maseda et al., 2014). Beside the mentioned cognitive performances, other cognitive parameters, such as naming ability, abstraction, verbal fluency and executive functioning can be used to predict dementia (Fabrigoule et al., 1998; Pusswald et al., 2013; Small, Herlitz, Fratiglioni, Almkvist, & Bäckmann, 1997).

Demographics and dementia

In addition to the advanced age of people with dementia, other demographic data, such as gender and education, must also be taken into account. Studies came to the conclusion that women have a higher prevalence rate to develop dementia than men. However, this could also be related to a higher life expectancy (Hugo & Ganguli, 2014; Lee, Lee, & Kim, 2017). Furthermore, a negative correlation between the duration of education and the development of dementia also applies (Caamaño-Isorna, Corral Montes Martinez, & Takkpuche, 2006; Ye et al., 2013).

Neuropsychological Test Battery Vienna

Early detection of dementia is very important, because timely intervention may offer the best chance of therapeutic success (Dubois et al., 2016; Pusswald et al., 2013). The goal of therapeutic interventions is to improve cognitions and everyday functioning, the delay of the ongoing disease progression and at the earliest stage of disease, the reduction of the conversion of MCI to the complete clinical picture of dementia (Lehrner et al., 2011). In order to detect dementia as early as possible, the NTBV has been developed. The NTBV assesses a broad range of cognitive abilities commonly affected by AD and other forms of dementia.

The assessment consists of different subtests, which in turn can be subdivided into different domains. These comprise the domains attention, language, executive functioning and memory. The NTBV can be applied to the patients in a short and an extended version, depending on the achieved test score of the previously performed Mini Mental State Examination (MMSE). The MMSE is a screening test for measuring cognitive function in elderly people. The test includes 30 questions and tasks. For each correct response, the patient scores one point. With a test score of 15-23 points, patients receive the short version of the NTBV. For a test result of 24 points or more, the extended version of the NTBV is used.

Objectivity of NTBV.

Objectivity, meaning the test results obtained with a test are independent of the test investigator, can be largely ensured by the detailed description of the instruction and the evaluation of the NTBV test scores (Kubinger, 2009).

Reliability of NTBV.

Reliability for NTBV can be seen as good for internal consistency and questionable to excellent for test-retest reliability. Reliability of a test describes the degree of accuracy to measure a particular mental characteristic. Internal consistency describes if items, which are supposed to measure the same construct, are coherent with one another and assess the same thing (Kubinger, 2009). Cronbach's alpha (α), which represents a value for internal consistency, for NTBV was conducted for different diagnostic groups with values ranging from 0.87 to 0.89 as well for the total sample with a high internal consistency (0.83 to 0.93; Hitzl, 2015; Lehrner et al., 2007; Macher, 2013). According to George and Mallery (2003) internal consistency is interpreted as excellent with $\alpha > 0.9$, good with $\alpha > 0.8$, acceptable with $\alpha > 0.7$, questionable with $\alpha > 0.6$, poor with $\alpha > 0.5$ and unacceptable with $\alpha < 0.5$. Another method to measure the degree of accuracy is the test-retest reliability. It estimates the stability of the items when measuring the same person on different examinations. To reach a good test-retest reliability, the same item given twice to the same person should correlate high. Consistency

over time was measured in recent studies with correlations ranging from 0.69 to 0.94 (Hitzl, 2015; Lehrner et al., 2007; Macher, 2013). A good test-retest reliability can be assumed at $r \geq 0.80$ with a time interval of four weeks (Chen et al., 2015, Gonçalves, Pinho & Simões, 2016).

Validity of NTBV.

Construct validity was determined by Moser (2018). It expresses if the test actually measures what it is supposed to measure and can be surveyed by factor analysis. The factor analysis is a method of identifying an amount of independent dimensions, sufficient to statistically explain a large number of correlating variable (Kubinger, 2009). Moser (2018) determined the dimensions by using principal component analysis and subsequent orthogonal Kaiser's Varimax rotation in six iterations. Only factors with intrinsic values higher than 1 were considered, which represent 65.3 % of the total variance. See table 1 for more information.

Table 1. *Rotated Component Matrix for NTBV subtests (n=1837) (Moser, 2018)*

	Components					
NTBV	1	2	3	4	5	6
TMT B	**-.82**	-.19	-.05	-.02	-.06	.10
TMT B - TMT A difference	-.78	-.14	-.05	-.04	-.05	.13
Stroop total /time	.74	.14	.10	.36	-.11	-.03
Digit-Symbol	.72	.26	.21	-.05	-.02	.05
5 Point Test	.71	.16	.11	-.07	.12	.23
Stroop difference	-.66	-.11	-.04	-.40	.15	.07
SWT	.65	.11	.32	-.02	.05	.01
PWT	.61	.06	.22	.04	.03	.02
Maze total/time	.46	.42	.15	-.09	.36	-.06
Concentration I	-.21	**-.82**	-.21	-.02	.01	-.03
TMT A	-.20	-.76	-.26	-.02	-.08	-.03
Concentration II	.42	.74	.24	-.01	.02	.02
C.I. Symbol	-.21	-.73	-.16	-.01	-.01	-.05
AKT	-.04	.63	.13	.08	.03	-.06
Interference C.I. total/time	.43	.56	.23	.21	.12	.09
BNT	.06	.45	.40	.00	.11	-.16
VSRT immediate recall	.28	.30	**.84**	.05	.01	.00
VSRT delayed recall	.27	.26	.83	.04	-.01	-.04
VSRT total recall	.21	.26	.80	.02	.00	-.00
VSRT recognition	.07	.21	.74	.03	.07	-.02
Stroop total	.16	.02	.01	**.76**	-.04	-.07
Interference C.I. total	-.11	.14	.11	.54	.50	.20
Planning Maze Test	.07	.05	.01	-.03	**.85**	-.10
5 Point Test Perseverations	-.01	-.01	-.07	-.02	-.06	**.93**
Intrinsic value (λ)	**5.03**	**3.88**	**3.27**	**1.23**	**1.19**	**1.06**
Determination Coefficient	**20.94%**	**16.18%**	**13.63%**	**5.14%**	**4.95%**	**4.42%**

There are six domains of the NTBV, which were examined empirically by means of cluster analysis based on healthy controls by Pusswald et al. (2013). Table 2 shows the six domains with the subtests and the corresponding R^2 with its own cluster and R^2 with the next closet cluster based on 250 cognitively healthy subjects (Pusswald et al., 2013).

Table 2. *A variable clustering procedure resulted in a six-cluster solution (Pusswald et al., 2013)*

Domain/Neuropsychological variable	R^2 with own cluster	R^2 with next closest
Domain 1/ Attention		
AKT	0.70	0.14
AKT total/time	0.69	0.14
TMT B	0.58	0.18
Digit-Symbol	0.44	0.24
TMT B -TMT A difference	0.43	0.07
C.I. Symbols	0.38	0.19
Domain 2/ Executive function-phonematic verbal fluency		
PWT	0.98	0.10
PWT l-words	0.72	0.08
PWT f-words	0.66	0.07
PWT b-words	0.64	0.09
Domain 3/ Executive function-interference		
Stroop colour words-words	0.87	0.17
Stroop total/time	0.86	0.16
Interference C. I. time	0.68	0.30
Interference C.I. total/time	0.67	0.26
Stroop colour words-colours difference	0.55	0.05
Stroop colour words-colour	0.55	0.24
Domain 4/ Language		
SWT total words	0.66	0.22
SWT supermarket	0.65	0.08
SWT animals	0.65	0.09
SWT tools	0.63	0.10
BNT	0.05	0.02
Domain 5/ Memory		
VSRT immediate recall	0.78	0.08
VSRT total recall	0.72	0.03
VSRT delayed recall	0.63	0.07
VSRT recognition	0.21	0.02
Domain 6/ Executive function-planning & nonverbal fluency		
Planning Maze Test	0.82	0.13
Planning Maze Test total/time	0.77	0.11
5 Point Test	0.45	0.14
TMT A	0.43	0.19

AKT Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT B* Trail Making Test Version B, *PWT* Phonematische Wortflüssigkeit, *SWT* Semantische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *TMT A* Trail Making Test Version A

Subtests of NTBV.

The domain attention is assessed by the Alters-Konzentrations-Test (AKT), a geriatric cancellation test (Gatterer, 1990). To complete the task, the test person is instructed to cross a determined sign in a row with other similar signs. The time needed to complete the task and the achieved score of the AKT are used to calculate the test score. The digit-symbol subtest of the German Wechsler Adult Intelligence Scale-Revised (Tewes, 1994) and the symbol counting task from the cerebral insufficiency test (C.I.; Lehrl & Fischer, 1997) are also used to determine attention. The digit-symbol test consists of numbers from one to nine, which are coded with symbols. Below the code is a row of the numbers one to nine repeated several times in random order. The task is to write as many symbols as possible under the corresponding number in 90 seconds. To complete the symbol counting task, the test person has to count all the squares represented with other symbols as fast as possible within 60 seconds. The Trail Making Test B (TMT B; Reitan, 1979) is a test where you have to connect letters and numbers in ascending order. The score difference of the Trail Making Test A (TMT A) and TMT B are also used to measure attention (Pusswald et al., 2013). Executive functioning is assessed in the NTBV by the subtests TMT A (Reitan, 1979), the Five Point Test (Regard, Strauss, & Knapp, 1982) and the Planning Maze Test from the Nürnberger Alters Inventar test battery (Oswald & Fleischmann, 1997). The TMT A consists of 25 numbers, which have to be connected with each other in ascending order. For the Five Point Test, the test person receives a sheet with several squares including five dots. The test person is asked to draw as many different figures as possible by connecting the dots with straight lines within three minutes. To draw a figure, not all dots have to be used. To complete the maze test, the test person has to pencil a way out of a labyrinth avoiding the blind alley. Also, the Stroop test from the Nürnberger Alters Inventar test battery (Oswald & Fleischmann, 1997) and the interference test from the C.I. (Lehrl & Fischer, 1997) are used to evaluate executive functioning. The Stroop test consists of two tasks. The first one shows different colours, which have to be named as fast as possible. The second task presents words written in different colours. The test person has to name the colours of the words, not the written words. The time needed, the right naming and the difference of both tests are used to calculate the final score. For the inference test, a row with several letters of A and B in different order is shown to the test person. The test person is instructed to name the letters in reverse order as fast as possible. The test score is calculated from the total score of right naming and the time needed to perform the test. The phonematic verbal fluency (PWT; Goodglass & Kaplan, 1983) are also used to assess executive function. The PWT includes three tasks, in which the test person has to name as many

words as possible in 60 seconds, starting with the letter b, f and l. Test score is obtained by counting the named words. Further the Boston naming Test (BNT; Morris et al., 1989) and semantic verbal fluency test (SWT; Goodglass & Kaplan, 1983) examine the domain language. To examine the test score of the BNT, pictures of different objects are shown to the test person, whose task is to name the objects. For the SWT, the test person is asked to name as many animals, grocery articles and tools as possible coming to her/his mind within 60 seconds. The sum of all named words forms the test score. The Verbal Selective Reminding Test (VSRT; Lehrner, Gleiß, Maly, Auff, & Dal-Bianco, 2006) with the subtests of immediate recall, total recall, delayed recall and recognition determine the domain memory (Crook, Salama, & Gobert, 1986; Youngjohn, Larrabee, & Crook, 1991). The task of the VSTR is to recall 15 words from a grocery list, which are shown to the test persons in five trials. Only those words which were not recalled by the test person are shown every trial after the first one. After 20 minutes, the test person has to recall all 15 words without seeing them again. The last task is a grocery list, which includes among other groceries the 15 previously presented words, which the test person has to recognize. More information about test procedure and test evaluation are described in the appendix.

Research question

Long before clinical symptoms of dementia are noticeable, neurodegenerative changes happen (Nestor, Scheltens, & Hodges, 2004). Therefore, early detection of dementia is necessary for appropriate intervention. It is well known, that timely countermeasures have an effective impact on the reduction of the disease symptoms (Hermann, Atri, & Salloway, 2017; Mao & Reddy, 2011). For example, the condition of MCI may stabilize or improve in many patients through neuropsychological interventions or psychotherapeutic measures (Lehrner et al., 2011). In addition, people are able to plan ahead, while they still have the cognitive ability to make important decisions for example about their future care or financial situation. Detecting dementia at an early stage, helps the patient and the family to gather information and advices about the disease to understand the behaviour of the affected one. To intervene as early as possible, valid and reliable instruments with a good diagnostic accuracy are necessary to detect dementia.

The aim of the study is to investigate if the NTBV is a valid predictor of neuropsychological testing in diagnosing dementia. Therefore, psychometric properties like internal consistency and test-retest reliability are used. Likewise, discrimination accuracy is investigated to gather more information on the diagnostic value of the NTBV. Additionally, reliable change analyses are used for different diagnostic groups to assess if a group has demonstrated

significant changes over time which cannot be accounted for by systematic or measurement error, like for example practice effects or missing test reliability. Differences between NTBV subtests across first and second examination and between NTBV subtests for converters and non-converters were expected. Likewise, differences for all subtests in diagnostic groups were expected except for comparison between HC and SCD groups. There should be no discrepancy between first and follow-up scores of NTBV in the same diagnostic group. A correlation between NTBV subtests and demographic variables was assumed.

Materials and methods

Participants

Data from 358 participants were analysed. The dataset was collected from 2008 to May 2018 within a current research project: the Vienna Conversion to Dementia Study. This study is a prospective cohort study encompassing consecutive, community-dwelling patients complaining of cognitive problems who were examined in the memory outpatient clinic for assessment of possible cognitive impairments and a medical referral. The same participants were surveyed during two measurements over 12 to 48 months. Before the testing, the participants received information about the questionnaires and an elaborated explanation of this study. They were also informed that the participation of the testing was voluntary. Furthermore, the participants were briefed that they would receive an invitation for a follow-up appointment 12 months later, which they could answer within a time frame of 12 to 48 months. The invitation refers to a voluntary follow-up study without any disadvantages, in case participants choose not to accept the invitation for the follow-up appointments. All data were collected anonymously and evaluated in a password-protected computer of the Medical University of Vienna – Department of Neurology, which can only be accessed by employees of the hospital. This ensures anonymity and confidentiality. The patients were allowed to stop the survey at any time. The surveys were based on a possible cognitive impairment and the referral to a doctor. All patients received standard laboratory blood test and clinical findings after having conducted the neuropsychological and cognitive testing. Magnetic resonance imaging (MRI) scan or a computer tomography (CT) scan of the brain was carried out for almost all patients. As a result, predominant advantages for the patients can be ensured. The neuropsychological assessment with the NTBV for each participant lasted approximately 60 minutes and was performed within one test session. For the purpose of the analysis, patients were divided into the subgroups healthy controls (HC), SCD, MCI and AD. SCD was diagnosed by the criteria of Jessen et al. (2014b) using an age, gender and education related mean score

greater than -1.50 SD in each domain of NTBV as cut off score. According to Petersen (2004), MCI was determined with a decline in cognition in at least one cognitive domain of NTBV below -1.50 *SD* of age-adjusted norm. AD was determined following the guidelines of DSM-5 criteria (APA, 2013) and NINCDS-ADRDA (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). A case example is demonstrated in figure 1 in the index of figures to illustrate the mentioned SD.

Exclusion criteria were applied for participants with a history of cerebral vascular pathology, severe head injury and the diagnosis AD at first examination. Also, current psychiatric diagnosis, any medical condition leading to severe cognitive impairments and an age under 51 led to the exclusion of participants. These diseases as well as impairments affect cognitive abilities and therefore have an influence on the study.

Neuropsychological measures

Only patients who performed the extended version of the NTBV at the first examination were considered in this study. The extended version was applied to all participants who achieved a MMSE score over 23. To assess IQ, Wortschatz-Test (WST IQ; Schmidt & Metzler, 1992), a standardized vocabulary test, was used. It is regularly applied to determine verbal comprehension in patients with dementia. For measuring depression, the patients filled out the Beck Depressions Inventory (BDI-II; Hautzinger, Keller & Kühner, 2006). The BDI-II has 21 questions asking about how often one has been feeling in certain ways for the past two weeks. Answers being scored on a four-point scale. Scores above 13 represent clinical depressive symptoms.

Statistical analyses

Statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 25.0. In order to control the type I error, Bonferroni correction was used. Because of the considerable difference of the sample size in diagnostic groups and a no interval scaling of the dependent variable, nonparametric methods were applied for all statistical analyses. Demographic and clinical characteristics are described by median and range due to a non-normal distribution. Cross-tabulation was performed to assess progression rates from first to second examination. To calculate the mean difference of NTBV score in all diagnostic groups, Kruskal-Wallis analysis was used. Post-hoc Bonferroni correction should show significant differences for all diagnosis comparisons except for HC and SCD. Next, one-tailed Friedman tests were performed for all diagnostic groups and for converters to AD and non-converters to explore the difference between the two examinations. Further, Mann-Whitney-U-Test was performed to assess the difference of the NTBV mean scores, demographic and clinical variables

between converters and non-converters. Following this, correlations between NTBV scores and demographic and clinical variables were conducted, using Spearman's rank correlation analysis. Internal consistency check for all items of the NTBV and different diagnostic groups were calculated using Cronbach's Alpha. In addition, test-retest reliability was assessed by Pearson correlation. To assess the predictive validity of NTBV scores for differentiating diagnostic groups, receiver operator characteristic (ROC) curves with AD as positive condition were used to determine sensitivity, specificity and cut-offs scores. These were chosen by the Youden Index. Sensitivity shows the probability that AD patients were correctly diagnosed with AD while specificity indicates the probability of correctly classified healthy patients. Positive and negative predicted values (PPV/NPV), positive and negative likelihood ratios (LR+/LR-) and percentage of correctly predicted patients were also calculated. PPV implies the probability of diagnosing AD when the patient has a positive test result whereas NPV indicates how probable a negative test rules out the disease. LR+ represents, how likely it is for someone to achieve a positive test if he has the disease compared with a person without the disease. LR- shows the opposite of LR+. The ROC curve demonstrates the discriminative ability of the NTBV, while the area under the curve (AUC) shows the accuracy of discrimination between converters to AD and non-converters (Eid, Gollwitzer, & Schmitt, 2013). Next, stepwise logistic regression was conducted to assess more detailed information about the discriminating ability of the NTBV among converters to AD and non-converters. Reliable Change Analyses were calculated with the Reliable Change Calculator, Version 1.0 (Hinton-Bayre, 2010). RCI determines the influence of disease progression over time and calculates the ability of each participant across both examinations taking into consideration changes not linked to cognitive deteriorations. This can be caused by practice effects or measurement errors. Mean test scores of the NTBV and their standard deviations for each diagnostic group and both examinations were used in this case to calculate the Reliable Change Index by Chelune et al. (1993) using the formula

$$RCI = ((X_2 - X_1) - (M_2 - M_1)) * SED$$

Thus, a minimum and maximum limit can be determined by calculating confidence intervals. The individual test score of a participant at both examinations is represented by X_2 and X_1, while M_2 and M_1 characterize the mean score of the comparing group at the two examinations. To calculate the standard error of difference (SED), SD from first examination and test-retest-reliability of the comparison HC group were used.

$$SED = \sqrt{2} * SEM$$

$$SEM = SD * \sqrt{(1 - test\text{-}retest\text{-}reliability)}$$

A significant change is shown at a change of ± 1.645.

The described test values from two-tailed test were recognized as statistically significant, using an alpha level of $p < 0.05$. To describe, if statistical significance has a meaningful magnitude, effect sizes were determined. Effect sizes basically demonstrate the strongness of an observed effect. To calculate the effect sizes *r*, asymptotic significance value (*z*) were divided by the root of the sample size (*n*).

$$r = \left| z / \sqrt{n} \right|$$

The reported effect sizes are interpreted as small effect with $r < 0.10$, as medium effect with $r < 0.30$ and as large effect with $r > 0.50$, according to Cohen (1992). Effect size *f* were determined using the formula

$$f = \sqrt{R^2 / 1 - R^2}$$

R^2 were calculated by logistic regression analysis. $f = 0.10$ represents a small effect, $f = 0.25$ corresponds to a medium effect and $f = 0.40$ indicates a large effect (Cohen, 1992).

Results

A total of 358 patients ranging from 51 to 94 years (*Mdn* = 70, *SD* = 9.12) were included in the study. The mean duration of the follow-up examination was 25.96 months (*SD* = 11.28) after the baseline examination. The interval for patients diagnosed with AD was the shortest (*M* = 23.15, *SD* = 9.93), while patients diagnosed with HC (*M* = 29.26, *SD* = 13.18) had the longest mean duration between first and second examination. SCD (*M* = 25.67, *SD* = 10.52) and MCI (*M* = 25.98, *SD* = 11.24) returned for the follow-up examination at about the same period of months. In the study 175 males (48.9 %) and 183 females (51.1%) were included. The duration of formal education was between six and 24 years (*Mdn* = 11.00, *SD* = 4.12). Table 3 shows baseline characteristics of the total sample and the different diagnostic groups.

Table 3. *Demographic and clinical characteristics of the sample*

	HC	SCD	MCI	Total
Age	61 (51-76)	66 (50-86)	70 (50-92)	69 (50-92)
Male/Female	30/16	21/21	124/146	175/183
Education	12 (6-24)	12 (8-22)	11 (8-22)	11 (6-24)
BDI-II	2 (0-29)	8 (0-31)	9 (0-50)	8 (0-50)
MMSE	29 (27-30)	29 (26-30)	28 (22-30)	28 (22-30)
WST-IQ	116 (90-139)	118 (88-133)	110 (77-140)	110 (77-140)

Note. Median (Range)

HC Healthy controls, *SCD* Subjective cognitive decline*, MCI* Mild cognitive impairment, *BDI-II* Beck-Depression-Inventory, *MMSE* Mini Mental State Examination, *WST-IQ* Wortschatztest Intelligenzquotient

At the baseline examination 270 patients were diagnosed as MCI, 46 as HC and 42 with SCD. At follow up 48 patients converted to AD from the total of 270 MCI patients. 183 patients were still diagnosed as MCI and 39 were diagnosed as SCD. All 46 HC remained with this diagnose. From the 42 SCD, 17 converted to MCI. The other 25 patients were still categorized as SCD. A description of the rates of conversion across both examinations for all diagnostic groups is shown in table 4.

Table 4. *Rates of conversion across examinations in the subgroups*

		Follow-up				
		HC	SCD	MCI	AD	Total
Baseline	HC	46 (100.0)	0	0	0	46
	SCD	0	25 (59.5)	17 (40.5)	0	42
	MCI	0	39 (14.4)	183 (67.8)	48 (17.8)	270
Total		46	64	200	48	358

Note. N (percent)

HC Healthy controls, *SCD S*ubjective cognitive decline*, MCI M*ild cognitive impairment, *AD* Alzheimer's disease

Kruskal-Wallis analysis showed significant differences for all NTBV subtests ($p \leq 0.01$) using diagnosis as grouping variables at first and second examination. Post-hoc Dunn Bonferroni analyses were conducted to correct for multiple testing and to reveal rates of com-

parison of each diagnostic group. At first examination no significant differences between HC and SCD group could be shown except for SWT. Looking at the second examination, there were no significant differences between diagnostic groups HC and SCD except for the NTBV scores Planning Maze Test ($p \leq 0.01$, $r = 0.30$) and Planning Maze Test total/time ($p < 0.05$, $r = 0.27$). The other group comparisons HC – MCI ($p < 0.05$), HC – AD ($p < 0.001$), SCD – MCI ($p < 0.05$) and SCD – AD ($p < 0.001$) confirmed significant mean differences of NTBV scores at both examinations. Most of the subtests for HC – SCD showed no effects (*r:* 0.00 to 0.32) at first examination and mainly small effects (*r*: 0.02 to 0.30) at second examination. Effect size for HC – MCI showed for the most subtests moderate effects at first examination (*r:* 0.15 to 0.52) and second examination (*r:* 0.18 to 0.53), whereas HC – AD comparisons revealed large effects from 0.48 to 1.14 at second examination. Effects for SCD – MCI ranged from 0.17 to 0.40 at baseline examination and from 0.18 to 0.39 at follow up examination, which indicates a small to medium effect size. SCD – AD comparison confirmed moderate to large effects (*r:* 0.40 to 0.96) at second examination. The remaining group comparison MCI – AD yielded significant differences ($p \leq 0.01$, *r:* 0.21 to 0.48) except for the NTBV scores C.I. Symbols ($p = 0.33$), PWT ($p = 0.19$) and Stroop colour words-colour ($p = 0.11$). Detail information is shown in table 5 and table 6.

Table 5. *Post-hoc Dunn Bonferroni comparison between diagnosis and NTBV scores with effect size (r) based on median at baseline examination. The reported effect sizes are interpreted as small effect with r < 0.10, as medium effect with r < 0.30 and as large effect with r > 0.50*

	HC-SCD	HC-MCI	SCD-MCI
AKT	r = 0.01	r = 0.32**	r = 0.31**
AKT total/time	r = 0.00	r = 0.33**	r = 0.32**
Digit-Symbol	r = 0.05	r = 0.35**	r = 0.30**
C.I. Symbols	r = 0.06	r = 0.25**	r = 0.20**
TMT A	r = 0.06	r = 0.33**	r = 0.28**
TMT B	r = 0.01	r = 0.33**	r = 0.32**
SWT	r = 0.32**	r = 0.52**	r = 0.28**
PWT	r = 0.12	r = 0.38**	r = 0.29**
BNT	r = 0.14	r = 0.30**	r = 0.19**
VSRT immediate recall	r = 0.25	r = 0.35**	r = 0.17**
VSRT total recall	r = 0.10	r = 0.38**	r = 0.31**
VSRT delayed recall	r = 0.08	r = 0.30**	r = 0.34**
VSRT recognition	r = 0.15	r = 0.15*	r = 0.25*
5 Point Test	r = 0.17	r = 0.38**	r = 0.25**
Stroop colour words-colour	r = 0.07	r = 0.29**	r = 0.31**
Stroop colour words-words	r = 0.10	r = 0.38**	r = 0.30**

Stroop total/time	r = 0.08	r = 0.25**	r = 0.30**
Stroop colour words difference	r = 0.21	r = 0.27**	r = 0.20**
Planning Maze Test	r = 0.02	r = 0.28**	r = 0.26**
Planning Maze Test total/time	r = 0.05	r = 0.28**	r = 0.24**
TMT B-TMT A difference	r = 0.03	r = 0.25**	r = 0.27**
Interference C. I. time	r = 0.05	r = 0.37**	r = 0.33**
Interference C.I. total/time	r = 0.04	r = 0.37**	r = 0.40**

** $p \leq 0.01$, * $p \leq 0.05$

HC Healthy controls, *SCD S*ubjective cognitive decline*, MCI* Mild cognitive impairment*, AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

Table 6. *Post-hoc Dunn Bonferroni comparison between diagnosis and NTBV scores with effect size (r) based on median at follow-up examination. The reported effect sizes are interpreted as small effect with r < 0.10, as medium effect with r < 0.30 and as large effect with r > 0.50*

	HC-SCD	HC-MCI	HC-AD	SCD-MCI	SCD-AD	MCI-AD
AKT	r = 0.17	r = 0.36**	r = 0.82**	r = 0.25**	r = 0.64**	r = 0.27**
AKT total/time	r = 0.16	r = 0.37**	r = 0.84**	r = 0.26**	r = 0.66**	r = 0.28**
Digit-Symbol	r = 0.10	r = 0.36**	r = 0.81**	r = 0.31**	r = 0.67**	r = 0.25**
C.I. Symbols	r = 0.22	r = 0.37**	r = 0.64**	r = 0.22**	r = 0.40**	r = 0.00
TMT A	r = 0.11	r = 0.33**	r = 0.80**	r = 0.27**	r = 0.68**	r = 0.29**
TMT B	r = 0.05	r = 0.41**	r = 0.96**	r = 0.40**	r = 0.85**	r = 0.31**
SWT	r = 0.21	r = 0.47**	r = 0.90**	r = 0.33**	r = 0.66**	r = 0.23**
PWT	r = 0.15	r = 0.53**	r = 0.71**	r = 0.32**	r = 0.53**	r = 0.15
BNT	r = 0.12	r = 0.31**	r = 0.70**	r = 0.24**	r = 0.58**	r = 0.24**
VSRT immediate recall	r = 0.07	r = 0.30**	r = 0.84**	r = 0.27**	r = 0.76**	r = 0.36**
VSRT total recall	r = 0.15	r = 0.41**	r = 1.14**	r = 0.31**	r = 0.96**	r = 0.48**
VSRT delayed recall	r = 0.14	r = 0.39**	r =1.07**	r = 0.31**	r = 0.91**	r = 0.44**
VSRT recognition	r = 0.14	r = 0.18*	r = 0.71**	r = 0.33**	r =0.91**	r = 0.43**
5 Point Test	r = 0.22	r = 0.35**	r = 0.73**	r = 0.25**	r = 0.56**	r = 0.21**
Stroop colour words-colour	r = 0.06	r = 0.33**	r = 0.64**	r = 0.32**	r = 0.54**	r = 0.16
Stroop colour words-words	r = 0.17	r = 0.38**	r = 0.85**	r = 0.28**	r = 0.66**	r = 0.26**
Stroop total/time	r = 0.22	r = 0.42**	r = 0.90**	r = 0.26**	r = 0.65**	r = 0.27**
Stroop colour words difference	r = 0.24	r = 0.28**	r = 0.98**	r = 0.18*	r = 0.58**	r = 0.29**
Planning Maze Test	r = 0.30**	r = 0.41**	r = 0.89**	r = 0.19**	r = 0.59**	r = 0.29**
Planning Maze Test total/time	r = 0.27*	r = 0.39**	r = 0.83**	r = 0.20**	r = 0.55**	r = 0.26**
TMT B-TMT A difference	r = 0.02	r = 0.37**	r = 0.88**	r = 0.39**	r = 0.82**	r = 0.29**
Interference C. I. time	r = 0.04	r = 0.31**	r = 0.48**	r = 0.32**	r = 0.72**	r = 0.29**
Interference C.I. total/time	r = 0.02	r = 0.29**	r = 0.76**	r = 0.31**	r = 0.73**	r = 0.30**

** $p \leq 0.01$, * $p \leq 0.05$

HC Healthy controls, *SCD S*ubjective cognitive decline*, MCI* Mild cognitive impairment, *AD* Alzheimer's disease*, AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

One-tailed Friedman tests analysing NTBV subtests score across baseline and follow-up examination for HC, SCD and MCI diagnosis group failed to show no significant differences for all subtests. The differences for HC are moderate to large (*r:* 0.30 to 0.68). Differences for SCD showed small effects ($r = 0.28$), while MCI yielded small to medium effects (*r:* 0.18 to 0.42). See table 1 in the list of tables for details.

Mann-Whitney-U-Tests were calculated to compare NTBV subtests scores and other relevant variables between converters to AD and non-converters. Significant differences were confirmed for all subtests, MMSE and age ($p < 0.001$) with small to moderate effect size (*r:* 0.17 to 0.48). No significant discrepancy was found for BDI-II ($p = 0.13$), IQ ($p = 0.69$), education ($p = 0.08$) and gender ($p = 0.87$). See table 7 for more details.

Spearman's rank correlations were conducted to compare the NTBV subtests with relevant sociodemographic variables for the total sample and subgroups. They show a significant correlation for the total sample with age ($p < 0.01$), BDI-II score ($p < 0.05$), WST IQ ($p < 0.05$) except for Interference C.I. total/time and education except for the subtest Stroop colour words – colour and Stroop colour words difference ($p < 0.01$). There are significant associations between some NTBV subtests and gender. For HC group, significant correlations could be found for some subtests with age, education and WST IQ ($p < 0.05$). BDI-II correlates only with two (TMT B, TMT B – TMT A difference) subtests and gender with four (PWT, VSRT immediate recall, VSRT total recall, Planning Maze Test total/time) subtests significantly. Looking at the SCD group, age, education and BDI-II showed significant correlations with subtests ($p < 0.05$), while gender correlate significantly with four (AKT total/time, VSRT delayed recall, VSRT recognition, 5 Point Test) and WST-IQ with three subtests (SWT, PWT, BNT). Significant correlations were determined for age and all subtests ($p < 0.05$) for MCI group. Education, BDI-II and WST IQ correlate significantly with some subtests. Significant correlations were conducted for gender with VSRT immediate recall, VSRT total recall, VSRT delayed recall and VSRT recognition. AD group showed only significant correlation for gender with Stroop colour words-words, for age with SWT, Planning Maze Test and Planning Maze Test total/time and for WST IQ with VSRT immediate recall subtests. No significant correlation could be found for age and education. The results are presented in Table 8a to 8e.

Table 7. *Mann-Whitney-U-Test between converters and non-converters based on median. Effect sizes are interpreted as small effect with r < 0.10, as medium effect with r < 0.30 and as large effect with r > 0.50*

	Non-converters	Converters	Cohen's d
AKT	33.00 ± 12.69	47.50 ± 18.24	0.33**
AKT total/time	1.62 ± 0.53	1.13 ± 0.39	0.33**
Digit-Symbol	43.00 ± 12.30	29.00 ± 10.58	0.31**
C.I. Symbols	20.00 ± 6.48	24.50 ± 8.18	0.24**
TMT A	40.00 ± 15.69	58.00 ± 29.54	0.29**
TMT B	96.00 ± 51.39	172.50 ± 76.48	0.37**
SWT	54.00 ± 15.36	39.00 ± 13.15	0.33**
PWT	31.00 ± 11.68	24.50 ± 11.21	0.17**
BNT	14.00 ± 1.21	14.00 ± 1.89	0.22**
VSRT immediate recall	8.00 ± 2.02	5.00 ± 1.63	0.38**
VSRT total recall	48.00 ± 10.17	31.00 ± 8.09	0.48**
VSRT delayed recall	10.00 ± 2.96	4.00 ± 2.81	0.48**
VSRT recognition	14.50 ± 1.50	11.50 ± 3.47	0.43**
5 Point Test	29.00 ± 9.82	18.00 ± 8.64	0.32**
Stroop colour words-colour	24.00 ± 5.75	29.50 ± 9.31	0.29**
Stroop colour words-words	46.00 ± 13.89	70.50 ± 24.23	0.35**
Stroop total/time	0.74 ± 0.22	0.49 ± 0.21	0.34**
Stroop colour words difference	23.00 ± 11.65	41.50 ± 18.12	0.33**
Planning Maze Test	37.00 ± 21.24	55.50 ± 25.64	0.31**
Planning Maze Test total/time	0.41 ± 0.19	0.26 ± 0.15	0.30**
TMTB-TMTA difference	55.50 ± 44.79	120.00 ± 60.94	0.35**
Interference C. I. time	22.00 ± 6.68	29.50 ± 10.37	0.32**
Interference C.I. total/time	1.55 ± 0.43	1.13 ± 0.34	0.33**
MMSE	28.50 ± 1.33	26.00 ± 1.83	0.39**
BDI-II	8.00 ± 8.02	6.50 ± 8.65	0.08
WST IQ	110.00 ± 12.82	110.00 ± 13.37	0.02
Education	12.00 ± 4.18	10.50 ± 3.63	0.09
Gender (male/ female)	48.7/ 51.3	50.0/ 50.0	0.01
Age	67.00 ± 8.90	75.00 ± 7.58	0.30**

Note. Median, Gender in percent, ** $p \leq 0.01$

AKT Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination, *BDI-II* Beck-Depressions-Inventory, *WST IQ* Wortschatztest Intelligenzquotient

Table 8a. *Spearman's rank correlation coefficient and Point-Biserial correlation coefficient between NTBV scores, MMSE and relevant variables for the total sample*

	Gender	Age	Education	BDI –II	WST IQ
AKT	0.08	0.49**	-0.23**	0.28**	-0.23**
AKT total/time	-0.07	-0.49**	0.24**	-0.28**	0.23**
Digit-Symbol	0.04	-0.50**	0.24**	-0.25**	0.31**
C.I. Symbols	0.11*	0.32**	-0.14**	0.21**	-0.20**
TMT A	0.07	0.50**	-0.12*	0.25**	-0.19**
TMT B	0.06	0.47**	-0.29**	0.26**	-0.31**
SWT	-0.05	-0.46**	0.24**	-0.29**	0.26**
PWT	0.08	-0.33**	0.25**	-0.28**	0.35**
BNT	-0.05	-0.33**	0.18**	-0.19**	0.26**
VSRT immediate recall	0.15**	-0.44**	0.20**	-0.13*	0.12*
VSRT total recall	0.14**	-0.56**	0.19**	-0.21**	0.20**
VSRT delayed recall	0.18**	-0.51**	0.18**	-0.18**	0.16**
VSRT recognition	0.14*	-0.35**	0.19**	-0.11*	0.14*
5 Point Test	-0.16**	-0.41**	0.29**	-0.23**	0.30**
Stroop colour words-colour	-0.05	0.33**	-0.09	0.16**	-0.25**
Stroop colour words-words	-0.04	0.48**	-0.25**	0.26**	-0.25**
Stroop total/time	-0.02	-0.50**	0.25**	-0.27**	0.25**
Stroop colour words difference	-0.05	0.45**	-0.31	0.20**	-0.17**
Planning Maze Test	0.17**	0.50**	-0.16**	0.26**	-0.19**
Planning Maze Test total/time	-0.18**	-0.46**	0.17**	-0.26**	0.19**
TMT B-TMT A difference	0.04	0.39**	-0.27**	0.23**	-0.29**
Interference C. I. time	0.07	0.39**	-0.24**	0.22**	-0.37**
Interference C.I. total/time	-0.06	-0.38**	0.23**	-0.22**	0.34

** $p \leq 0.01$, * $p \leq 0.05$

BDI - II Beck-Depressions.Inventory, *WST IQ* Wortschatztest Intelligenz quotient, *AKT* Alters Konzentrations Test, *C.I*. Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination

Table 8b**.** *Spearman's rank correlation coefficient and Point-Biserial correlation coefficient between NTBV scores and relevant variables for the HC group*

	Gender	Age	Education	BDI -II	WST IQ
AKT	-0.11	0.53**	-0.46**	0.16	-0.20
AKT total/time	0.12	-0.53**	0.48**	-0.17	0.19
Digit-Symbol	-0.03	-0.30*	0.40**	0.00	0.48**
C.I. Symbols	0.20	0.27	-0.23	0.16	-0.20
TMT A	-0.15	0.52**	-0.38*	0.23	-0.14
TMT B	-0.23	0.61**	-0.53**	0.41**	-0.42**
SWT	0.24	-0.44**	0.49**	-0.28	0-39**
PWT	0.46**	-0.22	0.37*	-0.21	0.17
BNT	0.28	-0.18	0.48**	-0.18	0.24
VSRT immediate recall	0.30*	-0.26	0.40**	0.03	0.12
VSRT total recall	0.30*	-0.38*	0.54**	-0.10	0.35*
VSRT delayed recall	0.25	-0.43**	0.57**	-0.22	0.44**
VSRT recognition	0.26	-0.35*	0.55**	-0.04	0.34*
5 Point Test	0.07	-0.37*	0.48**	0.30	0.45**
Stroop colour words-colour	-0.10	0.50**	-0.37*	0.19	-0.34*
Stroop colour words-words	-0.01	0.51**	-0.36*	0.13	-0.49**
Stroop total/time	-0.08	-0.45**	0.28	-0.23	0.39**
Stroop colour words difference	0.35	0.65**	-0.30	-0.20	-0.41
Planning Maze Test	0.29	-0.01	0.03	-0.02	-0.18
Planning Maze Test total/time	-0.36*	0.05	-0.08	0.05	0.19
TMT B-TMT A difference	-0.22	0.51**	-0.49**	0.41**	-0.48**
Interference C. I. time	0.09	0.54**	-0.46**	0.05	-0.38**
Interference C.I. total/time	0.06	-0.55**	0.39**	-0.14	0.31*

** $p \leq 0.01$, * $p \leq 0.05$

BDI - II Beck-Depressions.Inventory, *WST IQ* Wortschatztest Intelligenz quotient, *AKT* Alters Konzentrations Test, *C.I*. Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination

Table 8c. *Spearman's rank correlation coefficient and Point-Biserial correlation coefficient between NTBV scores and relevant variables for the SCD group*

	Gender	Age	Education	BDI -II	WST IQ
AKT	0.24	0.44**	-0.20	0.23	-0.04
AKT total/time	-0.26*	-0.44**	0.21	-0.23	0.05
Digit-Symbol	0.04	-0.45**	0.29	-0.28*	0.22
C.I. Symbols	0.22	0.36**	-0.17	-0.26*	0.10
TMT A	0.11	0.51**	-0.15	0.40**	-0.04
TMT B	-0.07	0.38**	-0.25	0.18	-0.10
SWT	-0.16	-0.27*	0.19	-0.25	0.31*
PWT	-0.04	-0.15	0.37**	-0.27*	0.34**
BNT	-0.13	-0.02	0.12	-0.18	0.28*
VSRT immediate recall	0.12	-0.41**	0.17	-0.11	0.15
VSRT total recall	0.21	-0.63**	0.24	-0.17	0.20
VSRT delayed recall	0.31*	-0.42**	0.12	-0.11	0.05
VSRT recognition	0.32**	-0.17	0.04	-0.04	0.11
5 Point Test	-0.31*	-0.43**	0.31*	-0.36**	0.08
Stroop colour words-colour	-0.12	0.21	0.01	0.05	-0.16
Stroop colour words-words	0.03	0.35**	-0.33**	0.25	-0.20
Stroop total/time	0.01	-0.35**	0.35**	-0.22	0.21
Stroop colour words difference	0.16	0.27*	-0.43**	0.30*	-0.24
Planning Maze Test	0.16	0.47**	-0.11	0.17	0.05
Planning Maze Test total/time	-0.16	-0.48**	0.18	-0.19	0.03
TMT B-TMT A difference	-0.14	0.22	-0.27*	0.00	-0.14
Interference C. I. time	0.13	0.27*	-0.25*	0.30*	-0.21
Interference C.I. total/time	-0.13	-0.25*	0.26*	-0.31*	0.23

** $p \leq 0.01$, * $p \leq 0.05$

BDI - II Beck-Depressions.Inventory, *WST IQ* Wortschatztest Intelligenz quotient, *AKT* Alters Konzentrations Test, *C.I*. Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination

Table 8d. *Spearman's rank correlation coefficient and Point-Biserial correlation coefficient between NTBV scores and relevant variables for the MCI group*

	Gender	Age	Education	BDI -II	WST IQ
AKT	0.03	0.38**	-0.19**	0.19*	-0.24**
AKT total/time	-0.02	-0.38**	0.19**	-0.19**	0.26**
Digit-Symbol	0.07	-0.44**	0.18*	-0.17*	0.26**
C.I. Symbols	0.02	0.18*	-0.15*	0.09	-0.31**
TMT A	0.06	0.40**	-0.05	0.13	-0.06
TMT B	0.12	0.35**	-0.22**	0.16*	-0.20**
SWT	-0.06	-0.36**	0.14	-0.12	0.21**
PWT	0.09	-0.19**	0.17*	-0.10	0.30**
BNT	-0.20	-0.29**	0.06	-0.11	0.08
VSRT immediate recall	0.20**	-0.34**	0.07	-0.03	-0.01
VSRT total recall	0.23**	-0.42**	0.02	-0.09	-0.02
VSRT delayed recall	0.31**	-0.38**	0.01	-0.07	-0.06
VSRT recognition	0.24**	-0.24**	0.06	-0.12	0.05
5 Point Test	-0.14	-0.31**	0.26**	-0.11	0.31**
Stroop colour words-colour	-0.02	0.24**	0.01	0.08	-0.14
Stroop colour words-words	-0.03	0.41**	-0.15*	0.11	-0.20**
Stroop total/time	0.01	-0.42**	0.15*	-0.11	0.20**
Stroop colour words difference	-0.04	0.40**	-0.18*	0.09	-0.17*
Planning Maze Test	0.13	0.38**	-0.16*	0.15*	-0.18*
Planning Maze Test total/time	-0.10	-0.35**	0.17*	-0.17*	0.17*
TMT B-TMT A difference	0.09	0.29**	-0.18*	0.14	-0.20**
Interference C. I. time	0.01	0.25**	-0.19**	0.17*	-0.33**
Interference C.I. total/time	-0.03	-0.25**	0.19**	-0.16*	0.32**

** $p \leq 0.01$, * $p \leq 0.05$

BDI - II Beck-Depressions Inventory, *WST IQ* Wortschatztest Intelligenz quotient, *AKT* Alters Konzentrations Test, *C.I*. Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination

Table 8e. *Spearman's rank correlation coefficient and Point-Biserial correlation coefficient between NTBV scores and relevant variables for the AD group*

	Gender	Age	Education	BDI -II	WST IQ
AKT	0.18	0.16	0.20	0.02	-0.43
AKT total/time	-0.14	-0.15	-0.25	0.05	0.42
Digit-Symbol	0.28	-0.02	-0.32	0.25	-0.02
C.I. Symbols	0.23	0.12	0.27	-0.03	-0.01
TMT A	0.14	0.20	0.16	0.02	-0.13
TMT B	0.00	0.07	0.31	-0.15	0.20
SWT	0.32	-0.42*	-0.16	-0.16	0.41
PWT	0.02	-0.36	-0.08	-0.13	0.21
BNT	-0.18	-0.06	0.29	-0.03	0.11
VSRT immediate recall	0.15	-0.16	0.10	0.16	0.50*
VSRT total recall	0.16	-0.23	-0.01	0.03	0.37
VSRT delayed recall	-0.19	-0.05	0.13	0.26	0.16
VSRT recognition	-0.16	-0.04	-0.01	0.14	0.31
5 Point Test	-0.23	0.06	-0.33	0.05	0.23
Stroop colour words-colour	-0.32	0.00	0.29	-0.04	0.07
Stroop colour words-words	-0.46*	0.30	0.15	0.13	0.14
Stroop total/time	0.37	-0.32	-0.10	-0.10	-0.10
Stroop colour words difference	-0.37	0.35	0.04	0.13	0.10
Planning Maze Test	0.28	0.43**	0.01	0.23	-0.25
Planning Maze Test total/time	-0.16	-0.31*	0.02	-0.33	0.29
TMT B-TMT A difference	0.03	-0.08	0.17	-0.12	0.26
Interference C. I. time	0.22	0.06	0.00	-0.12	-0.43
Interference C.I. total/time	-0.19	-0.08	0.01	0.14	0.44

** $p \leq 0.01$, * $p \leq 0.05$

BDI - II Beck-Depressions Inventory, *WST IQ* Wortschatztest Intelligenz quotient, *AKT* Alters Konzentrations Test, *C.I*. Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *MMSE* Mini Mental State Examination

Internal consistencies were computed for all NTBV scores and diagnostic groups using Spearman correlation to calculate Cronbach's Alpha. The diagnostic groups HC (α = 0.83), SCD (α = 0.87), MCI (α = 0.82) and the total sample (α = 0.86) revealed a high internal consistency ($\alpha > 0.80$). Cronbach's Alpha for AD achieved a value of 0.79. Pearson correla-

tion for different diagnostic groups failed to reach a good test-retest reliability (0.80 < *r*) except for NTBV scores of Digit-Symbol ($r = 0.81$) and VSRT recognition ($r = 0.87$) in SCD group and Stroop colour words – colour ($r = 0.85$) and Stroop colour words – words ($r = 0.81$) in HC group. Also, the Digit-Symbol subtest in the total sample reached a high test-retest reliability ($r = 0.82$). Test-retest reliability for NTBV scores ranged from 0.41 to 0.85 in HC group, 0.15 to 0.87 for SCD group, 0.39 to 0.80 for MCI group and 0.51 to 0.82 in the total sample group. A description of the test-retest reliabilities for the diagnostic groups is presented in table 9.

Table 9. *Test-retest reliability for subgroups (Pearson)*

	HC (N = 46)	SCD (N = 64)	MCI (N = 200)	Total sample (N = 358)
AKT	0.59**	0.62**	0.60**	0.64**
AKT total/time	0.68**	0.61**	0.66**	0.72**
Digit-Symbol	0.80**	0.81**	0.80**	0.82**
C.I. Symbols	0.41**	0.15	0.50**	0.51**
TMT A	0.61**	0.59**	0.55**	0.65**
TMT B	0.72**	0.66**	0.63**	0.71**
SWT	0.76**	0.62**	0.66**	0.76**
PWT	0.69**	0.73**	0.69**	0.75**
BNT	0.61**	0.42**	0.60**	0.67**
VSRT immediate recall	0.60**	0.31**	0.39**	0.52**
VSRT total recall	0.76**	0.67**	0.63**	0.78**
VSRT delayed recall	0.66**	0.53**	0.64**	0.73**
VSRT recognition	0.51**	0.87**	0.42**	0.58**
5 Point Test	0.47**	0.57**	0.66**	0.67**
Stroop colour words-colour	0.85**	0.67**	0.64**	0.72**
Stroop colour words-words	0.81**	0.73**	0.71**	0.75**
Stroop total/time	0.67**	0.74**	0.72**	0.75**
Stroop colour words difference	0.66**	0.51**	0.68**	0.69**
Planning Maze Test	0.65**	0.65**	0.51**	0.60**
Planning Maze Test total/time	0.61**	0.58**	0.66**	0.67**
TMT B-TMT A difference	0.67**	0.48**	0.56**	0.61**
Interference C. I. time	0.59**	0.73**	0.56**	0.64**
Interference C.I. total/time	0.77**	0.67**	0.61**	0.70**

** $p \leq 0.01$, * $p \leq 0.05$; *HC* Healthy controls, *SCD S*ubjective cognitive decline*, MCI* Mild cognitive impairment, *AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

ROC curve analysis was computed to assess the predictive power of the NTBV subtests. The scores of the NTBV subtests were used as predictors and patients who converted to AD were used as positive condition. Cut-off score were chosen by the Youden-Index. The cut-off score refers to a score that is assumed to represent the boundary between non-converters and converters to AD. Based on this cut-off score, sensitivity, specificity, PPV, NPV, LR+, LR- and percentage of correctly predicted patients were calculated. An LR+ above 10.00 indicates a large effect on increasing the probability of AD presence with a positive test result, an LR+ below 5.00 shows a small effect. For LR-, a value below 0.10 demonstrate a large effect on decreasing the probability of AD presence while a value above 0.50 reveals a small effect (Jaeschke, Guyatt, & Lijmer, 2002). The AUC was also calculated. An area of 0.50 represents a random guessing subtest for diagnosing AD, while an area over 0.90 shows an excellent diagnostic accuracy. For better understanding, two subtests are explained in detail. VSRT total recall reached a high sensitivity of 0.94 based on the cut-off score of 41.50. NPV can be seen as very high with a value of 0.99. LR- of 0.09 indicates a large effect. 74.0 percent were correctly classified at this cut-off score and the AUC score reached an excellent value of 0.91. Specificity, PPV and LR+ showed weak result for VSRT total recall. Looking at the subtest score Stroop colour words difference poor values were determined. At the cut-off score of 79.00 a very high sensitivity of 1.00 but in the same time a specificity of 0.00 could be shown. NPV and LR+ reached convincingly values but PPV and LR- failed to show good results. 14.7 percent of the sample were correctly classified with a diagnostic accuracy of 0.23. Detailed information about the other subtests are shown in table 2 in the list of tables.

The stepwise logistic regression model with NTBV subtests as predictor was defined significantly ($p < 0.001$) with $X^2 = 142.66$ and R^2 (Nagelkerke) = 0.64 for the entire sample with converters to AD as positive variable. The subtests VSRT delayed recall, VSRT total recall, VSRT recognition, TMT B, PWT and Stroop colour words – words yielded a significant *B*. The effect of R^2 by Nagelkerke indicates a strong effect with a value higher than $f = 0.40$ ($f = 1.33$; Cohen, 1992). The results are listed in table 10. Classification table of the stepwise logistic regression analysis shows 92.1% correctly predicted patients in the total sample. Thus, 97.0% of non-converters and 63.0 % of converters were correctly classified with the logistic regression model.

Table 10. *Parameters of stepwise logistic regression analysis with NTBV scores as predictors for converters vs. non-converters*

	Wald (1)	p	Exp(B)
VSRT delayed recall	5.95	0.02*	0.73
VSRT total recall	6.41	0.01 **	0.88
VSRT recognition	4.19	0.04*	0.83
TMT B	5.36	0.02*	0.19
PWT	8.98	$p < 0.001$**	1.07
Stroop colour words-words	5.08	0.02*	1.03

Note. Model characteristics: R^2 (Nagelkerke) = 0.64. Model X^2 = 142.66, significant $p < 0.001$

** $p \leq 0.01$, * $p \leq 0.05$

VSRT Verbal selective reminding test, *TMT B* Trail Making Test Version B, *PWT* Phonemmatische Wortflüssigkeit

RCI analyses were conducted to assess the performance of each participant in each diagnostic group for all NTBV scores across both examinations taking into consideration changes, which are not related to cognitive impairments. In order to clarify the calculations, a case example is explained in more detail. A patient from the AD group is compared to the HC group for the SWT subtest. First, mean score and SD of HC group is computed to calculate the confidence interval. Mean score of the SWT subtest at first examination was 75.15 with a SD of 13.94. At the follow-up examination mean score of the SWT subtest was 70.24 with a SD of 16.62 for HC group. Test-retest-reliability was assessed with a correlation of 0.69. The formula for SEM is applied by entering SD and test-retest reliability from HC group.

$$SEM = 13.94 * \sqrt{(1 - 0.69)} = 7.76$$

Now, SED is calculated by including SEM in the formula.

$$SED = \sqrt{2} * 7.76 = 3.94$$

A significant change can be assumed at a change of ± 1.645 (Chelune et al., 1993).

$$CI = 1.64 * 3.94 = 6.46$$

The CI is added or subtracted to the difference of M_2 and M_1 (DM). The DM is -4.91. The limit value of deterioration (-CI + DM) amount to -11.55, while the limit value of improvement (CI + DM) is 1.55. To demonstrate a statistically significant change between both examinations, RCI of the AD patient has to be higher or lower than the limit values. The patient chosen for the example has reached a test score of 30 ($X_{1)}$ at first examination and a test score of 26 (X_2) at second examination in the PWT subtest. RCI is calculated by including the mean score of the group and the individual patients score in the formula. For the PWT subtest the values are 0.69 for test-retest reliability, 5.77 for SEM, 3.40 for SED, 25.96 for M_1 and 24.48 for M_2.

$$RCI = ((26 - 30) - (24.48 - 25.96)) * 3.40 = -8.57$$

The RCI of -8.57 for this individual patient is compared with the limit values determined above. There is no higher or lower value than the upper or lower limiting value. Therefore, comparing to the HC group, no improvement or deterioration in the PWT subtest in this patient can be seen.

The percentage of participants, which showed a decline in NTBV subtests, ranged from 0 % to 18.8 % for SCD group. For MCI group, the deterioration in severe NTBV score ranged from 1.6% to 29.1 %. Converters to AD showed a decline between 0 % and 54.2% and non-converters deteriorated between 2.3 to 29.4 %. The percentage of SCD group revealed as improved ranged from 0 % to 73.4 % and from 0 % to 25.1 % for MCI group. For AD group, the improvement at second examination ranged from 0 % to 44.0 % and from 0.7 to 25.6 % for non-converters. The results of mean scores and SD of NTBV scores for HC group at first and second examination and the individual comparison between HC group and SCD group, MCI group, non-converters and converters are presented in table 3 and table 4 in the list of tables.

Discussion

The aim of the presented study was to evaluate the psychometric properties and the predictive power of the NTBV for early prediction of dementia. For this purpose, reliability and discrimination accuracy of the NTBV scores were examined for determining diagnostic groups. 40.5 % in the SCD group converted to MCI (17 of 42 cases) and 17.8 % of MCI group converted to AD (48 of 270 cases). This is similar to the results observed in prior investigations (Lehrner et al., 2005; Lehrner et al., 2016, Valencia & Lehrner, 2018). Some of the

patients, diagnosed with MCI at first examination, regressed to SCD group at second examination. Reasons can be lack of motivation or attention at first examination. No patient progressed from SCD directly to AD. Only patients with MCI converted to AD at second examination. This supports the assumption of the typical progression of dementia, starting from SCD to MCI and leading to AD in the course of time (Jessen et al., 2010; Jessen et al., 2014).

As expected, HC and SCD groups did not show significant differences between NTBV scores apart from the subtest SWT at first examination and Planning Maze test at second examination ($p = 0.01$). This underlines the criteria for SCD, showing normal age-, gender-, and education-adjusted performance on standardized cognitive test (Jessen et al., 2014, Mitchell et al., 2014; Sperling et al., 2011, McKhann et al., 2011). MCI and AD group did not differ in three subtests (C.I. Symbols, PWT, Stroop colour words – colour) at baseline examination, while all other group comparisons showed significant differences for all subtests at both examinations. The sample included patients with very different stages of dementia. Some of them, diagnosed with MCI or AD, had cognitive deterioration in only one cognitive domain while others suffered in various domains. Therefore, the reason for missing significant discrepancy between MCI and AD group in the three subtests mentioned could be highly similar scores. However, these findings underline the good discrimination power of the NTBV between diagnostic groups. Also, NTBV scores in the total sample group were significantly lower for most of the subtests at the follow-up examination representing the fundamental disease progression. Comparison between NTBV scores at first and second examination for SCD showed no differences except for the subtest BNT. This was not the case for HC and MCI, where many outliners were present. Here again, reasons could be aging, demotivation or lack of attention. These findings point to poor test-retest reliability and should be investigated in further researches.

For the domain attention, the predictive accuracy reached a good discrimination power except for one subtest. The AUC for the subtest AKT, Digit-Symbol-Test and C.I. Symbols were over 0.70 and for TMT B the AUC was 0.81. This means the subtest TMT B has a good diagnostic accuracy for converters to AD and non-converters. A recent study showed a higher AUC for AKT of 0.77 and for Digit-Symbol-Test of 0.87 (Lehrner et al., 2005). The TMT B – TMT A difference reached an AUC of 0.20, which means that the test is not appropriate to discriminate between diagnostic groups. The subtest of the language domain showed an AUC over 0.60 for BNT and an AUC of 0.78 for SWT. Therefore, only one subtest showed a fair discrimination power. Comparing the results to another study (Lehrner et al., 2007), which also examined ROC analysis, differences can be seen. In this study, all subtests of the lan-

guage domain reached an AUC between 0.79 to 0.87. Looking to the domain executive function, all subtest approached values over 0.70 for AUC except for the difference of Stroop colour words and PWT. Thus, all subtests, apart from the Stroop colour words difference and PWT, can be seen as fair to good predictors of future AD. These findings are consistent with other study results (Lehrner et al., 2007). The highest discrimination power with values over 0.80 for AUC was found in the domain memory. Here all subtests reached a very good to excellent diagnostic accuracy. These results of the memory domain agree with other studies investigating the predictive value of neuropsychological testing in detecting AD (Chen, Ratcliff, Belle, Cauley, DeKosky, & Ganguli, 2000; Lehrner et al., 2005, Lehrner et al., 2016). The results of the present study together with those of earlier studies support the assumption that the first affected cognition in dementia progress is the verbal memory.

Looking at the sensitivity of the NTBV, eight subtests (AKT, C.I. Symbols, TMT B, VSRT total recall, VSRT delayed recall, 5 Point Test, Planning Maze Test, Planning Maze total/time) reached a sensitivity over 0.80. This means that over 80.0% of patients were classified correctly as converters. Specificity for those eight subtests ranged from 0.50 to 0.78. This means, between 50.0% to 78.0% patients, depending on the subtest, were correctly categorized as not demented. A high sensitivity of a diagnostic test is important to avoid missing people who need the treatment. Regarding the specificity seven subtests (SWT, PWT, BNT, VSRT immediate recall, VSRT recognition, Stroop colour words-words, Stroop total/time) approached values over 0.80. Thus, the probability of being tested negative when the disease is absent is 80.0%. For those seven subtests the sensitivity ranges from 0.42 to 0.73. This means in only 42.0 to 73.0% of the cases, a person who has the disease will have a positive test result. The probability of not having AD if the test result is negative for the disease is high for all subtests ($0.90 \leq$ NPV), while the probability of having AD if the NTBV subtests show a positive result can be interpreted as weak (PPV $<$ 5). Concerning LR+, only one subtest (Stroop colour words-words) showed a high test result. For LR- barely three subtests (VSRT total recall, VSRT delayed recall, Planning Maze test) reached high to convincing values (Jaeschke et al., 1994). Thus, it may be concluded that the three subtests VSRT total recall, VSRT delayed recall and Planning Maze test have a high information value about how likely it is for someone to get a negative test if he or she does not have the disease, compared to a person with AD. This bears out the fact, that recall- and delayed-memory tasks are good predictors for detecting dementia (Chen et al., 2000; De Rotrou et al., 2007). The results of ROC analyses should be considered with caution since only 48 patients, developed from MCI to AD, were included as converters in the analyses. In many analyses poor values for discrimina-

tion accuracy were determined. One reason may be the given time interval between both examinations. The progression of dementia evolves over several years. The large range of 12 to 48 months can introduce different progressions extents regarding the severity of the disease. Also, in some cases the interval was not long enough to bring out the full extent of the pathology. Indeed, the sample consists of patients in very different stages of the disease. For example, patients diagnosed with MCI showed cognitive impairments in one, others in multiple domains of the NTBV. In addition, other diseases with similar symptoms, have been shown to affect the discrimination accuracy (Jaeschke et al., 1994, Muliyala & Varghese, 2010). Thus, stepwise logistic regression was used to survey the discrimination ability between converters to AD and non-converters of the NTBV subtest. Six subtests (VSRT delayed recall, VSRT total recall, VSRT recognition, TMT B, PWT, Stroop colour words-words) revealed to be significant. This means, all six subtests provide enough information to allow a correct classification for each patient. Regarding the six subtests, 92.1 % of patients in the total sample were diagnosed correctly, while only 7.9% received a wrong diagnosis. The results are consistent with the results of AUC, where all six subtests except for the PWT reached a good to excellent diagnostic accuracy.

Comparing the NTBV to the Vienna Visuo-Constructional Test 3.0 screening (VVT), another neuropsychological test to detect dementia, six subtests of the NTBV (VSRT immediate recall, VSRT delayed recall, VSRT total recall, VSRT recognition, TMT B, Stroop colour words-words) reached a higher AUC score than the VVT (0.80). Most of the subtests of the NTBV had a higher score for sensitivity ($0.6 < SE$) while only three subtests showed the same or higher specificity ($0.86 \leq SP$). PPV of 0.63 for VVT is higher than every subtest while NPV of VVT (NPV = 0.84) is lower than all subtest of NTBV. Comparing LR+ and LR- between VVT and NTBV, no meaningful differences can be reported. To sum it up, VVT has a higher ability of correctly detecting non-converters and a higher informative value of the probability having AD, when the test results are positive. In contrast, the NTBV shows a higher ability of correctly diagnosing converters to AD and a higher validity of the probability that a person with a negative test result for AD truly do not have the disease. Future researches could focus on combining both neuropsychological tests to improve the diagnostic accuracy.

RCI analysis were conducted to predict cognitive outcome of NTBV subtests for each patient. Looking at the result of SCD group, deterioration was under 10.0% for most of the subtests of NTBV. Regarding the improvement of SCD group, positive changes were not higher than 10.0% for most of the subtests. An exception is the BNT with 37.5% and the In-

terference C.I. time with 73.4 % improvement. Comparing the MCI group, no differences between first and second examination were expected. Further, the reliable change index for the MCI group, the percentage of patients exhibiting significant decline, was up to 29.1%. Improvements in this diagnostic group could also be shown. Deterioration for non-converters were lower than 10 % for half of the subtests. BNT and Stroop colour words-words reached deterioration over 20%. Improvements were mostly under 10% except for SWT, BNT, Stroop colour words-words, Stroop colour words difference and TMT B – TMT A difference. For converters to AD, a deterioration higher than 10% in all subtests except for six subtests (SWT, PWT, 5 Point Test, Planning Maze Test total/time, TMT B -TMT A difference, Interference C.I. total/time) was observed. Deterioration for this diagnostic group ranged up to 54.2%. These results underline the cognitive decline of patients suffering from the progression of dementia. Nevertheless, there were also unexpected apparent improvements in AD group for the subtests AKT, BNT, Stroop colour words-colour, Stroop colour words-words, Stroop colour word difference and TMT B -TMT A difference. The explanation for the significant positive change in all diagnostic groups are mainly practice effects. NTBV was conducted twice with the same sample. In related samples improvement in test performance can be reached only by practice effects (Foki et al., 2017). The reason for the high deterioration could be lack of motivation or error effects. In addition, aging, misclassification and different states of disease progression in diagnostic group can have an influence on the test results. The unusual cognitive changes in all diagnostic groups could be another explanation why the diagnostic accuracy for some subtests of the NTBV could not reach convincing values.

Total internal consistency can be reported as strong with a Cronbach's α of 0.86. The subgroups HC (r = 0.83), SCD (r = 0.87) and MCI (r = 0.82) reached a good reliability as well. Internal consistency for AD (r = 0.79) can be interpreted as acceptable. According to George and Mallery (2003) internal consistency can be seen as good with Cronbach's $\alpha > 0.80$ and acceptable with Cronbach's $\alpha > 0.70$. From the aspect of internal consistency, it may be concluded that the NTBV is a reliable measurement. The test-retest reliability was inappropriate in most of the subtest for all diagnostic groups. Acceptable reliability values for standard neuropsychological measurements of demented people are generally considered at $r \geq 0.80$ with a time interval of four weeks between both examinations (Chen et al., 2015, Gonçalves, Pinho & Simões, 2016). The values of this study are generally located below this value with test-retest reliability as low as r = 0.31. This means, it is questionable, whether the values provided by the NTBV are trustable and if the scores are valid representations of the patients' performance. The test-retest reliability reached a good value for the subtest Digit-

Symbol for all diagnostic groups ($r > 0.80$) and as well for the VSRT recognition in the SCD group ($r = 0.87$) and for Stroop colour words-colour ($r = 0.85$) and Stroop colour words-words ($r = 0.81$) in the HC group. Any other tests showed an unacceptable to acceptable reliability over time (*r:* 0.31 to 0.77). González-Blanch et al. (2010) showed similar result for the subtest TMT A ($r = 0.61$) and TMT B ($r = 0.69$) in HC after a time interval of three months. Regarding test-retest-reliability in VSRT subtests, values of 0.64 for VSRT immediate recall, 0.82 for VSRT total recall and 0.67 for VSRT delayed recall could be shown in a recent study with the same time interval of this study (Lehrner et al., 2006). For BNT subtest, Katsumata et al. (2015) were able to find a higher test-retest reliability of 0.74 for MCI group than in this study. This should be interpreted carefully because no information about the time interval between both examinations could be found for Katsumata et al.'s study (2015). Comparing the test results to other studies, again one reason for the poor test-retest reliability could be the long time between the first and second examination. Mostly, reliability will be higher when the interval between both examinations is rather short. A previous study has come to this conclusion by examining the VSRT subtests (Lehrner et al., 2006). In addition, aging can have an influence on test results. Further studies should include a younger patients sample to investigate the influence of the age and using the same time interval to enable a more accurate comparison.

Correlation between NTBV subtests with relevant sociodemographic variables showed significant association with age, gender, BDI-II score, WST IQ score and education ($p < 0.05$). It can be concluded, that lower test scores were in line with higher age and depression symptoms and vice versa. In the same manner, the results show that higher IQ and education leads to a better score in the different subtests and vice versa. Positive and negative correlation for different subtests can be explained by mixed rating scales. The results are consistent with previous studies. For this reason, there are age-, gender- and education-adjusted norms. Also, severity of depression is considered while making the diagnoses (Lehrner et al., 2014; Meng & D'Arcy, 2012; Sharp & Gatz, 2011). As there were no significant differences for BDI-II, WST IQ, education and gender between converters and non-converters, effect of these variables can be excluded.

While interpreting the results of the study, some limitations should be taken into account. Patients with a MMSE score under 24 at second examination received the short version of the NTBV because they were not capable of filling in the long version of the test. In consequence there are missing values, which leads to a reduction of the representativeness of the sample. In addition, the results cannot be extended to the general population due to the selec-

tive sample, which only consisted of people who were referred from a doctor or actively made an appointment because of cognitive complaints. Moreover, there is no information about cognitive training or other treatments to maintain the cognitive functions and delay the progression to AD between the interval of both examinations (Carrière, 2009; Hill, Mowszowski, Naismith, Chadwick, Valenzuela, & Lampit, 2016; Ngandu et al., 2015). Also, current mood, which may influence cognitive performance while testing and personality, which has an impact on cognitive complaints were not investigated in this study (Boone, 2009; Jessen et al., 2014, Marino et al., 2009). Cognitive training or treatment evaluation, current mood and personality should be considered in future research. Despite the limitations, one of the strengths of the study is the quite large sample size, which covered a range from healthy to variously ill persons. Further, all patients received a thorough neuropsychological examination. Therefore, detailed data and accurate diagnoses were provided.

Measurement instruments play an important role in clinical practice. Neuropsychological tests are gold standard in diagnosing cognitive dysfunction related to dementia. Knowing about the disease at an early stage can help to delay cognitive deterioration and maintain patients' independence. Therefore, the quality of the instrument is essential. In conclusion, the study demonstrates that the NTBV can be used as predictor of neuropsychological testing in diagnosing dementia but because of questionable psychometric criteria for some subtests, the diagnosis should be interpreted carefully. Future research should focus on subtest, which had a good predictive value for diagnosing AD. This can increase the predictive power and guarantee the overall test efficiency.

References

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Association.

Baudic, S., Dalla Barba, G., Thibaudet, M. C., Smagghe, A., Remy, P., & Traykov, L. (2006). Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology*, *21*(1), 15-21.

Bäckman, L., Small, B. J., & Fratiglioni, L. (2001). Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain*, *124*(1), 96-102.

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2016). *Neuroscience: exploring the brain.* Philadelphia: Wolters Kluwer.

Bennett, R. G. (1995). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Journal of the American Geriatrics Society, 43*(4), 460. https://doi.org/10.1111/j.1532-5415.1995.tb05832.x

Berr, C., Wancata, J., & Ritchie, K. (2005). Prevalence of dementia in the elderly in Europe. *European Neuropsychopharmacology*, *15*(4), 463-471.

Boone, K.B. (2009). Fixed belief in cognitive dysfunction despite normal neuropsychological scores: Neurocognitive hypochondriasis?. *Clinical Neuropsychology, 23*, 1016–1036.

Caamaño-Isorna, F., Corral, M., Montes-Martínez, A., & Takkouche, B. (2006). Education and dementia: a meta-analytic study. *Neuroepidemiology*, *26*(4), 226-232.

Carrière, I., Fourrier-Reglat, A., Dartigues, J. F., Rouaud, O., Pasquier, F., Ritchie, K., & Ancelin, M. L. (2009). Drugs with anticholinergic properties, cognitive decline, and dementia in an elderly general population: the 3-city study. *Archives of internal medicine*, *169*(14), 1317-1324.

Chelune, G.J., Naugle, R.I., Lueders, H., Sedlak, J., & Awad, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology 7*(1), 41–52.

Chen, C. W., Chu, H., Tsai, C. F., Yang, H. L., Tsai, J. C., Chung, M. H., ... & Chou, K. R. (2015). The reliability, validity, sensitivity, specificity and predictive values of the Chinese version of the Rowland Universal Dementia Assessment Scale. *Journal of clinical nursing*, *24*(21-22), 3118-3128.

Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*, *55*(12), 1847-1853.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 122*(1), 155-159.

Crook, T., Salama, M., & Gobert, J. (1986). A computerized test battery for detecting and assessing memory disorders. *Senile dementias: Early detection*, 79-85.

De Rotrou, J., Battal-Merlet, L., Wenisch, E., Chausson, C., Bizet, E., Dray, F., ... & Hanon, O. (2007). Relevance of 10-min delayed recall in dementia screening. *European journal of neurology*, *14*(2), 144-149.

Drummond, C., Coutinho, G., Fonseca, R. P., Assunção, N., Teldeschi, A., de Oliveira-Souza, R., ... & Mattos, P. (2015). Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. *Frontiers in aging neuroscience*, *7*, 1-11.

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., … & Jack, C. R. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's and Dementia, 12*(3), 292-323. https://doi.org/10.1016/j.jalz.2016.02.002

Eid, M., Gollwitzer, M., & Schmitt, M. (2013). *Statistik und Forschungsmethoden: Lehrbuch.* Basel: Beltz Verlag.

Elias, M. F., Beiser, A., Wolf, P. A., Au, R., White, R. F., & D'Agostino, R. B. (2000). The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Archives of Neurology*, *57*(6), 808–813. https://doi.org/10.1001/archneur.57.6.808

Fabrigoule, C., Rouch, I., Taberly, A., Letenneur, L., Commenges, D., Mazaux, J. M., … & Dartigues, J. F. (1998). Cognitive process in preclinical phase of dementia. *Brain*, *121*(1), 135–141. https://doi.org/10.1093/brain/121.1.135

Foki, T., Hitzl, D., Pirker, W., Novak, K., Pusswald, G., Auff, E., & Lehrner, J. (2017). Assessment of individual cognitive changes after deep brain stimulation surgery in Parkinson's disease using the Neuropsychological Test Battery Vienna short version. *Wiener Klinische Wochenschrift*, *129*(15-16), 564-571.

Gatterer, G. (1990). *Alters-Konzentrations-Test (AKT)*. Göttingen: Hogrefe.

George, D. & Mallery, M. (2003). *Using SPSS for Windows step by step: a simple guide and reference 11.0 update.* Boston: Allyn & Bacon.

Glodzik-Sobanska, L., Reisberg, B., De Santi, S., Babb, J. S., Pirraglia, E., Rich, K. E., … & De Leon, M. J. (2007). Subjective memory complaints: Presence, severity and future outcome in normal older subjects. *Dementia and Geriatric Cognitive Disorders*, *24*(3), 177–184. https://doi.org/10.1159/000105604

Gonçalves, M. M., Pinho, M. S., & Simões, M. R. (2016). Test–retest reliability analysis of

the Cambridge Neuropsychological Automated Tests for the assessment of dementia in older people living in retirement homes. *Applied Neuropsychology: Adult*, *23*(4), 251-263.

González-Blanch, C., Pérez-Iglesias, R., Rodríguez-Sánchez, J. M., Pardo-García, G., Martínez-García, O., Vázquez-Barquero, J. L., & Crespo-Facorro, B. (2010). A digit symbol coding task as a screening instrument for cognitive impairment in first-episode psychosis. *Archives of Clinical Neuropsychology*, *26*(1), 48-58.

Goodglass, H. & Kaplan, P. (1983). *The Assessment of Aphasia and Related Disorders*. Philadelphia: Lea & Fabinger.

Grober, E., Lipton, R. B., Hall, C., & Crystal, H. (2000). Memory impairment on free and cued selective reminding predicts dementia. *Neurology*, *54*(4), 827–832. https://doi.org/10.1212/WNL.54.4.827

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depressions-Inventar (BDI-II)*. Frankfurt: Harcourt Test Services.

Herman, L., Atri, A., & Salloway, S. (2017). Alzheimer's Disease in Primary Care: The Significance of Early Detection, Diagnosis, and Intervention. *The American journal of medicine*, *130*(6), 756.

Hill, N. T., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2016). Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. *American Journal of Psychiatry*, *174*(4), 329-340.

Hinton-Bayre, A. D. (2010). Deriving Reliable Change Statistics from Test–Retest Normative Data: Comparison of Models and Mathematical Expressions. *Archives of Clinical Neuropsychology, 25(3)*, 244-256.

Hitzl, D. (2013). *Neuropsychologischer Outcome nach tiefer Hirnstimulation bei Parkinsonpatienten – eine retrospektive Datenauswertung.* Unveröffentliche Diplomarbeit, Medizinische Universität Wien.

Höfler, S., Bengough, T., Winkler, P., & Griebler, R. (2015). Österreichischer Demenzbericht 2014. *Bundesministerium für Gesundheit und Sozialministerium, Wien*.

Hugo, J. & Ganguli, M. (2014). Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine*, *30*(3), 421-442.

Jacobs, D. M., Sano, M., Dooneief, G., Marder, K., Bell, K. L., & Stern, Y. (1995). Neuropsychological detection and characterization of preclinical Alzheimer's disease. *Neurology*, *45*(5), 957–962. https://doi.org/10.1212/WNL.45.5.957

Jaeschke, R., Guyatt, G., & Lijmer, J. (2002). Diagnostic tests. *Users' Guides to the Medical Literature–A Manual for Evidence-Based Clinical Practice. AMA Press, Chicago*, 121-140.

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Brill-Edwards, P., ... & Haynes, B. (1994). Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test B. What Are the Results and Will They Help Me in Caring for My Patients?. *Jama*, *271*(9), 703-707.

Jessen, F., Amariglio, R. E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., … & Wagner, M. (2014b). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's and Dementia, 10*(6), 844-852. https://doi.org/10.1016/j.jalz.2014.01.001

Jessen, F., Wiese, B., Bachmann, C., Eifflaender-Gorfer, S., Haller, F., Kölsch, H., … & Bickel, H. (2010). Prediction of dementia by subjective memory impairment effects of severity and temporal association with cognitive impairment. *Archives of General Psychiatry*, *67*(4), 414–422. https://doi.org/10.1001/archgenpsychiatry.2010.30

Jessen, F., Wolfsgruber, S., Wiese, B., Bickel, H., Mösch, E., Kaduszkiewicz, H., … & Wagner, M. (2014a). AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimer's and Dementia, 10*(1), 76-83. https://doi.org/10.1016/j.jalz.2012.09.017

Kaindl, M. & Schipfer, R. K. (2017). Familien in Zahlen 2016. *Österreichisches Institut für Familienforschung: Vienna, Austria*.

Katsumata, Y., Mathews, M., Abner, E. L., Jicha, G. A., Caban-Holt, A., Smith, C. D., ... & Fardo, D. W. (2015). Assessing the discriminant ability, reliability, and comparability of multiple short forms of the Boston naming test in an Alzheimer's disease center cohort. *Dementia and geriatric cognitive disorders*, *39*(3-4), 215-227.

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Vienna: Hogrefe Verlag.

Kumar, A. & Foster, T.C. (2007). Neurophysiology of Old Neurons and Synapses. In D.R., Riddle (Ed.). *Brain Aging: Models, Methods, and Mechanisms*. Boca Raton (FL): CRC Press/Taylor & Francis.

Lee, D. Y., Lee, J. H., Ju, Y. S., Kang Uk Lee, M. D., Kim, K. W., Jhoo, J. H., ... & Woo, J. I. (2002). The prevalence of dementia in older people in an urban population of Korea: the Seoul study. *Journal of the American Geriatrics Society*, *50*(7), 1233-1239.

Lehrl, S. & Fischer, B. (1997). *Kurztest für Cerebale Insuffizienz (c.I.-Test)*. Ebersberg: Vless.

Lehrner, J., Bodendorfer, B., Lamm, C., Moser, D., Dal-Bianco, P., Auff, E., & Pusswald, G. (2016). Subjective memory complaints and conversion to dementia in patients with subjective cognitive decline and patients with mild cognitive impairment. *Zeitschrift für Neuropsychologie*, *27*(2), 85 – 93.

Lehrner, J., Bodner, T., Dal-Bianco, P., & Schmidt, R. (2011). Demenzsyndrome. In J., Lehrner, G., Pusswald, E., Fertl, W., Strubreither, & I. Kryspin-Exner (Ed.). *Klinische Neuropsychologie* (pp. 375-394). Vienna: Springer.

Lehrner, J., Gleiß, A., Maly, J., Auff, E., & Dal-Bianco, P. (2006). Der Verbale Selektive Reminding Test (VSRT). Ein Verfahren zur Überprüfung verbaler Gedächtnisfunktionen. *Neuropsychiatrie*, *20*(3), 204–214.

Lehrner, J., Gufler, R., Guttmann, G., Maly, J., Gleiß, A., Auff, E., & Dal-Bianco, P. (2005). Annual conversion to Alzheimer disease among patients with memory complaints attending an outpatient memory clinic: The influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *Wiener Klinische Wochenschrift*, *117*(18), 629–635. https://doi.org/10.1007/s00508-005-0428-6

Lehrner, J., Maly, J., Gleiß, A., Auff, E. & Dal-Bianco, P. (2007). The Vienna Neuropsychological Test Battery (VNTB) for detecting dementia: standardization, norms, and validation. *Psychologie in Österreich, 4*, 358–365.

Lehrner, J., Maly, J., Gleiss, A., Auff, E., & Dal-Bianco, P. (2008). Neuropsychological performance and mild cognitive impairment subtypes in patients reporting cognitive problems attending a memory outpatient clinic. *European Journal of Geriatrics*, *10*(2), 59-68.

Lehrner, J., Moser, D., Klug, S., Gleiß, A., Auff, E., Dal-Bianco, P., & Pusswald, G. (2014). Subjective memory complaints, depressive symptoms and cognition in patients attending a memory outpatient clinic. *International Psychogeriatrics*, *26*(3), 463-473.

Luck, T., Luppa, M., Matschinger, H., Jessen, F., Angermeyer, M. C., & Riedel-Heller, S. G. (2015). Incident subjective memory complaints and the risk of subsequent dementia. *Acta Psychiatrica Scandinavica*, *131*(4), 290–296. https://doi.org/10.1111/acps.12328

Macher, S. (2013). *Reliabilität und Validität der Neuropsychologischen Testbatterie Vienna bei Patienten mit leichter Demenz – eine Pilotstudie.* Unveröffentlichte Diplomarbeit, Medizinische Universität Wien.

Mao, P. & Reddy, P. H. (2011). Aging and amyloid beta-induced oxidative DNA damage and mitochondrial dysfunction in Alzheimer's disease: implications for early intervention and therapeutics. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1812*(11), 1359-1370.

Marino, S. E., Meador, K. J., Loring, D. W., Okun, M. S., Fernandez, H. H., Fessler, A. J., ... & Schoenberg, M. R. (2009). Subjective perception of cognition is related to mood and not performance. *Epilepsy & Behavior*, *14*(3), 459-464.

Maseda, A., Lodeiro-Fernández, L., Lorenzo-López, L., Núñez-Naveira, L., Balo, A., & Millán-Calenti, J. C. (2014). Verbal fluency, naming and verbal comprehension: three aspects of language as predictors of cognitive impairment. *Aging & mental health*, *18*(8), 1037-1045.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939–939. https://doi.org/10.1212/WNL.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., ... & Mohs, R. C. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 263-269.

Meng, X. & D'Arcy, C. (2012). Education and dementia in the context of the cognitive reserve hypothesis: a systematic review with meta-analyses and qualitative analyses. *PloS one*, *7*(6), e38268.

Mitchell, A. J., Beaumont, H., Ferguson, D., Yadegarfar, M., & Stubbs, B. (2014). Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: Meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(6), 439–451. https://doi.org/10.1111/acps.12336

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., Van Belle, G., Fillenbaum, G. D. M. E., ... & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*(9), 1159-1165.

Moser, S. (2018). Cognitive functions and mortality in patients with subjective cognitive decline, mild cognitive impairment and alzheimer's disease. Unveröffentliche Diplomarbeit, Medizinische Universität Wien.

Muliyala, K. P. & Varghese, M. (2010). The complex relationship between depression and dementia. *Annals of Indian Academy of Neurology*, *13*(2), 69-73.

Nestor, P.J., Scheltens P., & Hodges J.R. (2004). Advances in the early detection of Alzheimer's disease. *Nature Reviews Neuroscience, 10,* 34-41.

Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., ... & Lindström, J. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*, *385*(9984), 2255-2263.

Oswald, W. D. & Fleischmann, U. M. (1997). *Das Nünberger-Alters-Inventar*. Göttingen: Hogrefe Verlag.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256*(3), 183-194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*(9), 1133–1142. https://doi.org/10.1212/WNL.56.9.1133

Pusswald, G., Moser, D., Gleiß, A., Janzek-Hawlat, S., Auff, E., Dal-Bianco, P., & Lehrner, J. (2013). Prevalence of mild cognitive impairment subtypes in patients attending a memory outpatient clinic-comparison of two modes of mild cognitive impairment classification. Results of the vienna conversion to dementia study. *Alzheimer's and Dementia*, *9*(4), 366–376. https://doi.org/10.1016/j.jalz.2011.12.009

Reitan, R. (1979). *Trail Making Test (TMT)*. Tucson: Reitan Neuropsychology Laboratory

Regard, M., Strauss, E., & Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Perceptual and Motor Skills*, *55*(3), 839–844. https://doi.org/10.2466/pms.1982.55.3.839

Reisberg, B., Shulman, M. B., Torossian, C., Leng, L., & Zhu, W. (2010). Outcome over seven years of healthy adults with and without subjective cognitive impairment. *Alzheimer's and Dementia*, *6*(1), 11–24. https://doi.org/10.1016/j.jalz.2009.10.002

Reynish, E., Bickel, H., Fratiglioni, L., Kiejna, A., Prince, M., & Georges, J. (2009). Systematic review and collaborative analysis of the prevalence of dementia in Europe. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *5*(4), 393.

Scheichenberger, S. (2018). Epidemiologie und Einteilung der Demenzformen. In S., Scheichenberger & B., Scharb (Ed.). *Spezielle validierende Pflege* (4th ed., pp. 47). Berlin, Heidelberg: Springer.

Schiffczyk, C., Romero, B., Jonas, C., Lahmeyer, C., Müller, F., & Riepe, M. W. (2013). Efficacy of short-term inpatient rehabilitation for dementia patients and caregivers: pro-

spective cohort study. *Dementia and geriatric cognitive disorders*, *35*(5-6), 300-312.

Schmidt, K. H., & Metzler, P. (1992). *Wortschatztest: WST*. Beltz.

Sharp, E. S. & Gatz, M. (2011). Relationship between education and dementia: an updated systematic review. *Alzheimer disease and associated disorders*, *25*(4), 289-304.

Small, B. J., Herlitz, A., Fratiglioni, L., Almkvist, O., & Bäckman, L. (1997). Cognitive predictors of incident Alzheimer's disease: a prospective longitudinal study. *Neuropsychology*, *11*(3), 413–420. https://doi.org/MCI-Converted #6

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., ... & Park, D. C. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 280-292.

Tewes, U. (1994). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision 1991 (HAWIE-R).* Bern: Hans Huber.

Thal, D. R. & Braak, H. (2005). Postmortale Diagnosestellung bei Morbus Alzheimer: Stadiengliederungen der kennzeichnenden Hirnveränderungen. *Der Pathologe, 26*(3), 201-213. https://doi.org/10.1007/s00292-004-0695-4

Valencia, N. & Lehrner, J. (2018). Screening for dementia with the Vienna Visuo-Constructional Test 3.0 screening (VVT 3.0 screening). *Neuropsychiatrie*, *32*(4), 196-203.

Wittchen, H. U. & Hoyer, J. (2011). *Klinische Psychologie & Psychotherapie*. Heidelberg: Springer.

Ye, B. S., Seo, S. W., Cho, H., Kim, S. Y., Lee, J. S., Kim, E. J., ... & Park, K. W. (2013). Effects of education on the progression of early-versus late-stage mild cognitive impairment. *International psychogeriatrics, 25*(4), 597-606.

Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1991). First-last names and the grocery list selective reminding test: Two computerized measures of everyday verbal learning. *Archives of Clinical Neuropsychology*, *6*(4), 287–300. https://doi.org/10.1016/0887-6177(91)90006-U

Index of figures

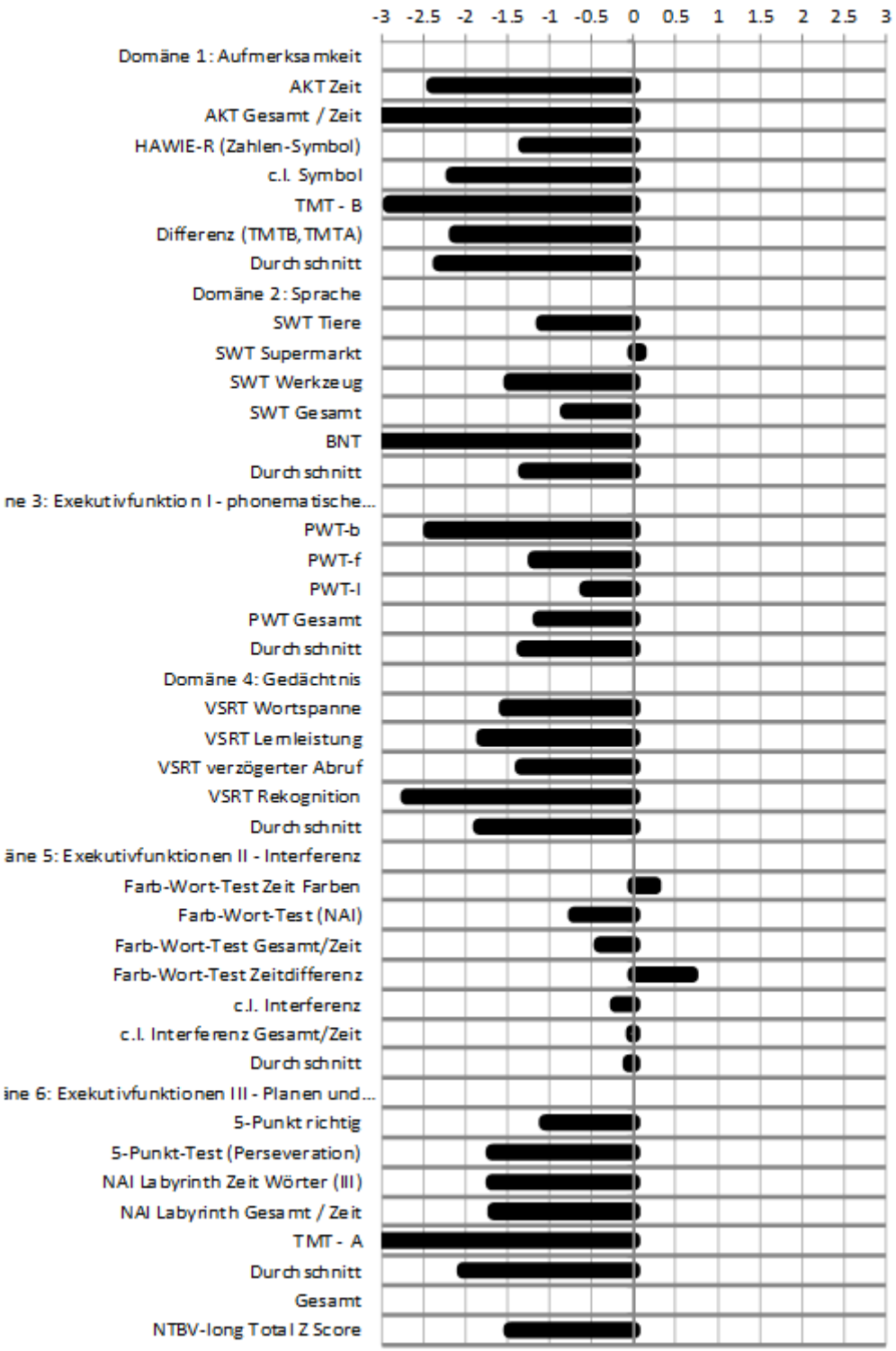

Figure 1. Age-, gender- and education-adjusted standard deviation of mean scores of a case example for each subtest. In this study SCD was diagnosed using a mean score greater than 1.50 SD in each domain of NTBV. MCI was determined with a decline in cognition in at least one cognitive domain of NTBV by -1.50 *SD* below the norm.

Table 1. *One-tailed Friedman test comparing NTBV scores across examination 1 & 2 for HC, SCD & MCI group using Bonferroni correction. The reported effect sizes are interpreted as small effects with r < 0.10, as medium effect with r < 0.30 and as large effect with r > 0.50.*

	HC			SCD			MCI		
	Baseline	Follow-up	Cohen's r	Baseline	Follow-up	Cohen's r	Baseline	Follow-up	Cohen's r
AKT	26.50 ± 11.00	27.50 ± 8.51	0.14	28.00 ± 9.10	30.50 ± 9.43	0.20	36.00 ± 12.98	36.00 ± 14.92	0.05
AKT total/time	2.06 ± 0.54	2.08 ± 0.55	0.12	1.91 ± 0.55	1.76 ± 0.53	0.08	1.50 ± 0.46	1.50 ± 0.49	0.02
Digit-Symbol	50.50 ± 11.88	48.50 ± 11.41	0.30*	47.50 ± 10.24	48.00 ± 9.32	0.16	38.00 ± 11.51	36.00 ± 11.45	0.21**
C.I. Symbols	18.00 ± 3.55	17.00 ± 4.10	0.10	18.50 ± 7.44	20.00 ± 5.12	0.17	21.00 ± 6.44	23.00 ± 7.64	0.28**
TMT A	32.50 ± 10.93	31.00 ± 10.93	0.11	36.50 ± 11.90	36.00 ± 10.71	0.08	44.00 ± 16.47	46.00 ± 21.17	0.11
TMT B	74.50 ± 38.67	69.50 ± 45.17	0.09	80.00 ± 29.54	79.00 ± 24.72	0.16	107.00 ± 54.81	118.00 ± 65.91	0.31**
SWT	76.50 ± 13.94	74.00 ± 16.62	0.44**	57.50 ± 12.69	60.00 ± 9.32	0.01	50.00 ± 12.45	48.00 ± 14.01	0.13
PWT	40.50 ± 11.63	41.00 ± 10.74	0.01	35.00 ± 9.82	36.00 ± 8.83	0.23	28.00 ± 10.46	28.00 ± 11.76	0.08
BNT	15.00 ± 0.58	15.00 ± 0.58	0.00	14.00 ± 0.81	15.00 ± 0.69	0.28*	14.00 ± 1.37	14.00 ± 1.25	0.03
VSRT immediate recall	9.00 ± 1.87	9.00 ± 2.37	0.37**	8.00 ± 1.83	8.00 ± 2.28	0.12	7.00 ± 1.91	6.00 ± 1.97	0.27**
VSRT total recall	55.00 ± 9.08	55.50 ± 9.48	0.01	51.50 ± 9.23	52.00 ± 9.63	0.18	45.00 ± 9.41	41.00 ± 9.93	0.37**
VSRT delayed recall	12.00 ± 2.74	12.00 ± 2.38	0.30*	11.00 ± 2.64	11,00 ± 2.90	0.16	9.00 ± 2.87	8.00 ± 3.81	0.42**
VSRT recognition	14.75 ± 0.93	15.00 ± 2.19	0.21	15.00 ± 1.17	15.00 ± 1.14	0.10	14.50 ± 1.67	14.00 ± 2.60	0.39**
5 Point Test	35.00 ± 8.62	35.50 ± 10.33	0.15	32.00 ± 8.77	32.00 ± 8.48	0.04	25.00 ± 9.24	26.00 ± 8.65	0.00
Stroop colour words-colour	21.00 ± 4.80	21.50 ± 3.92	0.07	22.00 ± 3.95	22.00 ± 4.43	0.17	25.00 ± 5.96	26.00 ± 7.01	0.24**
Stroop colour words-words	37.00 ± 9.55	39.50 ± 9.70	0.05	42.00 ± 9.89	44.00 ± 9.31	0.07	50.00 ± 14.36	52.00 ± 18.26	0.13
Stroop total/time	0.84 ± 0.22	0.92 ± 0.24	0.68**	0.84 ± 0.19	0.79 ± 0.22	0.08	0.69 ± 0.21	0.69 ± 0.23	0.08
Stroop colour words difference	16.00 ± 6.51	17.00 ± 7.23	0.13	20.00 ± 8.32	21.00 ± 9.75	0.13	25.00 ± 12.40	25.00 ± 14.51	0.05
Planning Maze Test	29.50 ± 14.60	26.00 ± 11.88	0.58**	35.50 ± 15.44	34.50 ± 12.59	0.03	40.00 ± 23.18	40.00 ± 24.80	0.10
Planning Maze Test total/time	0.50 ± 0.22	0.58 ± 0.28	0.41**	0.43 ± 0.18	0.44 ± 0.16	0.01	0.38 ± 0.18	0.36 ± 0.19	0.10
TMT B-TMT A difference	38.50 ± 34.58	36.50 ± 40.26	0.17	43.50 ± 23.85	45.00 ± 19.48	0.12	62.00 ± 49.10	72.50 ± 56.44	0.31**
Interference C. I. time	19.00 ± 5.01	20.00 ± 6.17	0.25	20.00 ± 4.19	20.00 ± 4.62	0.12	24.00 ± 7.00	25.00 ± 8.23	0.20
Interference C.I. total/time	1.79 ± 0.41	1.70 ± 0.49	0.33*	1.67 ± 0.36	1.70 ± 0.32	0.07	1.42 ± 0.41	1.36 ± 0.40	0.18

Note. Median and Standard deviation, ** $p \leq 0.01$, * $p \leq 0.05$

HC Healthy controls, *SCD* Subjective cognitive decline, *MCI* Mild cognitive impairment, *AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

Table 2. *Results of analyses of sensitivity, specificity, positive and negative predicted values and likelihood ratios, percent correctly predicted at the chosen cut-off value, and receiver operating characteristics (AUC) with 95% confidence intervals*

	Cut off	SE	SP	PPV	NPV	LR+	LR-	%	AUC
AKT (N_{con} = 48; N_{non} = 310)	36.50	0.83 (0.70-0.93)	0.61 (0.55-0.66)	0.25 (0.22-0.29)	0.96 (0.93-0.98)	2.13 (1.77-2.58)	0.27 (0.14-0.52)	64.0 (58.8-69.0)	0.78 (0.72-0.85)
AKT T/T (N_{con} = 48; N_{non} = 310)	1.38	0.75 (0.60-0.86)	0.71 /0.66-0.76)	0.29 (0.24-0.34)	0.95 (0.92-0.97)	2.58 (2.03-3.28)	0.35 (0.21-0.58)	71.5 (66.5-76.1)	0.78 (9.72-0.85)
Digit (N_{con} = 48; N_{non} = 310)	36.50	0.75 (0.60-0.86)	0.67 (0.62-0.72)	0.26 (0.22-0.31)	0.95 (0.91-0.97)	2.28 (1.81-2.86)	0.37 (0.23-0.61)	68.2 (63.1-73.0)	0.76 (0.69-0.84)
C.I. Symbols (N_{con} = 48; N_{non} = 309)	19.63	0.83 (0.70-0.91)	0.50 (0.44-0.56)	0.21 (0.18-0.23)	0.95 (0.91-0.97)	1.66 (1.40-1.97)	0.33 (0.18-0.64)	54.3 (49.0-59.6)	0.70 (0.62-0.78)
TMT A (N_{con} = 48; N_{non} = 310)	53.50	0.58 (0.43-0.72)	0.79 (0.74-0.83)	0.30 (0.24-0.37)	0.92 (0.90-0.95)	2.78 (2.02-3.84)	0.53 (0.38-0.74)	76.3 (71.5-80.6)	0.75 (0.68-0.82)
TMT B (N_{con} = 48; N_{non} = 310)	113.50	0.83 (0.70-0.93)	69.27 (64.21–74.02)	69.27 (64.21–74.02)	69.27 (64.21–74.02)	69.27 (64.21–74.02)	69.27 (64.21–74.02)	69.3 (64.2–74.0)	0.81 (0.74-0.88)
SWT (N_{con} = 48; N_{non} = 310)	42.50	0.63 (0.47-0.76)	0.81 (0.76-0.85)	0.34 (0.27-0.41)	0.93 (0.91-0.95)	3.28 (2.39-4.51)	0.46 (0.32-0.67)	78.5 (73.9-82-6)	0.78 (0.71-0.85)
PWT (N_{con} = 48; N_{non} = 309)	21.50	0.42 (0.28-0.57)	0.81 (0.76-0.85)	0.25 (0.18-0.34)	0.90 (0.87-0.92)	2.18 (1.45-3.27)	0.72 (0.56-0.92)	75.63(70.8-80.0)	0.64 (0.56-0.73)
BNT (N_{con} = 48; N_{non} = 309)	13.50	0.46 (0.31-0.61)	0.80 (0.75-0.84)	0.26 (0.20-0.34)	0.90 (0.88-0.93)	2.28 (1.56-3.34)	0.68 (0.52-0.88)	75.4 (70.5-79.7)	0.68 (0.59-0.76)
VSRT IR (N_{con} = 48; N_{non} = 310)	5.50	0.60 (0.45-0.74)	0.86 (0.82-0.90)	0.40 (0.32-0.49)	0.93 (0.91-0.95)	4.36 (3.04-6.24)	0.46 (0.32-0.65)	82.7 (78.4-86.5)	0.82 (0.75-0.88)
VSRT TR (N_{con} = 48; N_{non} = 310)	41.50	0.94 (0.83-0.99)	0.71 (0.66-0.76)	0.33 (0.29-0.38)	0.99 (0.96-1.00)	3.23 (2.67-3.90)	0.09 (0.03-0.26)	74.0 (69.2-78.5)	0.91 (0.87-0.95)
VSRT DR (N_{con} = 48; N_{non} = 309)	7.50	0.90 (0.77-0.97)	0.78 (0.73-0.82)	0.39 (0.33-0.44)	0.98 (0.95-0.99)	4.07 (3.23-5.13)	0.13 (0.06-0.31)	79.6 (75.0-83.6)	0.90 (0.85-0.95)
VSRT rec (N_{con} = 48; N_{non} = 309)	13.23	0.73 (0.58-0.85)	0.85 (0.81-0.89)	0.43 (0.36-0.51)	0.95 (0.93-0.97)	4.90 (3.57-6.73)	0.32 (0.20-0.51)	83.7 (79.4-87.3)	0.85 (0.78-0.91)
5 Point Test (N_{con} = 48; N_{non} = 310)	25.50	0.85 (0.72-0.94)	0.59 (0.53-0.65)	0.24 (0.21-0.28)	0.96 (0.93-0.98)	2.08 (1.75-2.49)	0.25 (0.12-0.49)	62.6 (57.3-67.6)	0.77 (0.70-0.85)
Stroop C (N_{con} = 48; N_{non} = 309)	25.50	0.79 (0.65-0.90)	0.64 (0.58-0.69)	0.25 (0.22-0.30)	0.95 (0.92-0.97)	2.20 (1.79-2.71)	0.33 (0.19-0.57)	66.1 (60.9-71.0)	0.75 (0.67-0.83)
Stroop W (N_{con} = 48; N_{non} = 308)	63.50	0.63 (0.47-0.76)	0.88 (0.84-0.91)	0.45 (0.36-0.54)	0.94 (0.91-0.96)	5.20 (3.58-7.56)	0.43 (0.30-0.62)	84.6 (80.4-88.1)	0.80 (0.73-0.87)
Stroop T/T (N_{con} = 47, N_{non} = 301)	0.54	0.62 (0.46-0.75)	0.86 (0.82-0.90)	0.41 (0.33-0.50)	0.94 (0.91-0.95)	4.42 (3.09-6.34)	0.45 (0.31-0.64)	82.8 (78.4-86.6)	0.79 (0.71-0.86)
Stroop dif (N_{con} = 48; N_{non} = 378)	79.00	1.00 (0.93-1.00)	0.00 (0.00-0.01)	0.15 (0.15-0.15)	100.00 (-)	1.00 (-)	0.00 (-)	14.7 (11.1-19.0)	0.23 (0.15-0.31)
Maze Test (N_{con} = 48; N_{non} = 309)	41.50	0.88 (0.75-0.95)	0.62 (0.56-0.68)	0.26 (0.23-0.30)	0.97 (0.94-0.99)	2.31 (1.93-2.76)	0.20 (0.09-0.43)	65.6 (60.4-70.5)	0.76 (0.70-0.83)
Maze T/T (N_{con} = 48; N_{non} = 309)	0.36	0.85 (0.72-0.94)	0.61 (0.55-0.66)	0.25 (0.22-0.29)	0.96 (0.93-0.98)	2.18 (1.82-2.62)	0.24 (0.12-0.48)	64.2 (58.9-69.1)	0.75 (0.68-0.82)

	Cut off	SE	SP	PPV	NPV	LR+	LR-	%	AUC
TMT dif (N_{con} = 47; N_{non} = 310)	36.50	0.83 (0.70-0.93)	0.61 (0.55-0.66)	0.25 (0.22-0.29)	0.96 (0.93-0.98)	2.13 (1.77-2.58)	0.27 (0.14-0.52)	64.0 (58.8-69.0)	0.78 (0.72-0.85)
C.I. Inter (N_{con} = 48; N_{non} = 319)	1.38	0.75 (0.60-0.86)	0.71 (0.66-0.76)	0.29 (0.24-0.34)	0.95 (0.92-0.97)	2.58 (2.03-3.28)	0.35 (0.21-0.58)	71.5 (66.5-76.1)	0.78 (9.72-0.85)
C.I. Inter T/T (N_{con} = 48; N_{non} = 303)	36.50	0.75 (0.60-0.86)	0.67 (0.62-0.72)	0.26 (0.22-0.31)	0.95 (0.91-0.97)	2.28 (1.81-2.86)	0.37 (0.23-0.61)	68.2 (63.1-73.0)	0.76 (0.69-0.84)

SE Sensitivity, *SP* Specificity, *PPV* Positive predicted value, *NPV* Negative predicted value, *LR+* Positive likelihood ratio, *LR-* Negative likelihood ratio, *%* percent correctly predicted, *AUC* = Area under curve, N_{con} converters, N_{non} non-converters, *AKT* Alters Konzentrations Test, *T/T* total/time, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test, *IR* immediate recall, *TR* total recall, *DR* delayed recall, *rec* recognition, *C* colour, *W* words, *dif* difference, *TMT dif* TMT B – TMT A difference, *Inter* Interference

Table 3. *Results of individual comparison between HC group and SCD/ MCI group*

	HC			SCD			MCI		
Test	Baseline	Follow-up	rtt	% Deteriorated	% No change	% Improved	% Deteriorated	% No change	% Improved
AKT	29.96 ± 11.01	28.33 ± 8.51	0.59	1.6	92.2	6.2	4.0	86.0	10.0
AKT total/time	2.00 ± 0.54	2.07 ± 0.55	0.68	7.8	90.6	1.6	5.0	93.5	1.5
Digit-Symbol	51.07 ± 11.88	48.65 ± 11.41	0.80	4.8	93.6	1.6	5.2	91.2	3.6
C.I. Symbols	18.41 ± 3.55	18.09 ± 4.10	0.41	9.4	81.2	9.4	8.6	86.3	5.1
TMT A	33.59 ± 10.93	34.65 ± 10.93	0.61	4.7	89.0	6.3	15.6	74.3	10.1
TMT B	83.59 ± 38.67	81.33 ± 45.17	0.72	1.6	93.6	4.8	24.6	68.6	6.8
SWT	75.15 ± 13.94	70.24 ± 16.62	0.76	3.2	82.5	14.3	2.6	85.8	11.6
PWT	41.85 ± 11.63	41.76 ± 10.74	0.69	4.8	93.6	1.6	1.6	91.0	7.4
BNT	14.72 ± 0.58	14.72 ± 0.58	0.61	18.8	43.7	37.5	29.1	45.8	25.1
VSRT immediate recall	9.20 ± 1.87	8.48 ± 2.37	0.60	6.3	81.2	12.5	10.1	84.4	5.5
VSRT total recall	55.46 ± 9.08	55.30 ± 9.48	0.76	10.9	82.9	6.2	16.6	78.9	4.5
VSRT delayed recall	11.22 ± 2.75	11.89 ± 2.39	0.66	10.9	86.0	3.1	19.1	78.4	2.5
VSRT recognition	14.41 ± 0.93	13.92 ± 2.19	0.51	0.0	100.0	0.0	19.6	74.4	6.0
5 Point Test	35.41 ± 8.62	34.70 ± 10.33	0.47	3.2	93.6	3.2	1.6	94.2	4.2
Stroop colour words-colour	21.65 ± 4.80	21.85 ± 3.92	0.85	7.9	82.6	9.5	19.3	68.9	11.8
Stroop colour words-words	39.59 ± 9.55	39.59 ± 9.70	0.81	9.5	81.0	9.5	14.4	74.4	11.2
Stroop total/time	0.84 ± 0.22	0.98 ± 0.24	0.67	12.9	87.1	0.0	11.7	88.3	0.0
Stroop colour words difference	16.65 ± 6.51	15.59 ± 7.23	0.66	7.9	82.6	9.5	15.2	74.7	16.8
Planning Maze Test	33.61 ± 14.60	27.98 ± 11.88	0.65	7,8	87.5	4.7	18.6	71.3	10.1
Planning Maze Test total/time	0.54 ± 0.22	0.64 ± 0.28	0.61	7.8	92.2	0.0	5.1	93.9	1.0
TMT B-TMT A difference	50.02 ± 34.58	46.67 ± 40.26	0.67	3.2	95.2	1.6	6.3	72.1	21.6
Interference C. I. time	19.48 ± 5.01	20.70 ± 6.17	0.59	1.6	25.0	73.4	7.5	86.0	6.5
Interference C.I. total/time	1.83 ± 0.41	1.72 ± 0.49	0.77	3.1	90.7	6.2	5.8	89.5	4.7

Note. Mean and Standard deviation

HC Healthy controls, *SCD* Subjective cognitive decline, *MCI* Mild cognitive impairment, *rtt* Test-retest reliability, *AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wortflüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

Table 4. *Results of individual comparison between HC group and non-converters/converters*

	HC			Non-converters			Converters		
Test	Baseline	Follow-up	rtt	% Deteriorated	% No change	% Improved	% Deteriorated	% No change	% Improved
AKT	29.96 ± 11.01	28.33 ± 8.51	0.59	3.5	88.4	8.1	10.9	65.0	24.1
AKT total/time	2.00 ± 0.54	2.07 ± 0.55	0.68	5.2	92.2	2.6	11.1	88.9	0.0
Digit-Symbol	51.07 ± 11.88	48.65 ± 11.41	0.80	5.6	91.1	3.3	11.5	77.0	11.5
C.I. Symbols	18.41 ± 3.55	18.09 ± 4.10	0.41	11.4	81.8	6.8	20.0	68.9	11.1
TMT A	33.59 ± 10.93	34.65 ± 10.93	0.61	12.0	78.9	9.1	34.8	52.2	13.0
TMT B	83.59 ± 38.67	81.33 ± 45.17	0.72	16.3	77.4	6.3	54.2	33.3	12.5
SWT	75.15 ± 13.94	70.24 ± 16.62	0.76	3.0	84.6	12.4	0.0	96.0	4.0
PWT	41.85 ± 11.63	41.76 ± 10.74	0.69	2.3	90.7	7.0	0.0	96.0	4.0
BNT	14.72 ± 0.58	14.72 ± 0.58	0.61	23.9	50.5	25.6	50.0	23.9	26.1
VSRT immediate recall	9.20 ± 1.87	8.48 ± 2.37	0.60	8.4	84.8	6.8	26.1	73.9	0.0
VSRT total recall	55.46 ± 9.08	55.30 ± 9.48	0.76	8.7	84.5	6.8	17.4	76.1	6.5
VSRT delayed recall	11.22 ± 2.75	11.89 ± 2.39	0.66	14.6	82.8	2.6	26.1	71.7	2.2
VSRT recognition	14.41 ± 0.93	13.92 ± 2.19	0.51	10.4	84.4	5.2	52.2	34.8	13.0
5 Point Test	35.41 ± 8.62	34.70 ± 10.33	0.47	2.7	92.9	4.4	0.0	92.0	8.0
Stroop colour words-colour	21.65 ± 4.80	21.85 ± 3.92	0.85	15.5	75.4	9.1	20.0	52.0	28.0
Stroop colour words-words	39.59 ± 9.55	39.59 ± 9.70	0.81	29.4	48.6	22.0	32.0	44.0	24.0
Stroop total/time	0.84 ± 0.22	0.98 ± 0.24	0.67	11.1	88.2	0.7	20.0	80.0	0.0
Stroop colour words difference	16.65 ± 6.51	15.59 ± 7.23	0.66	11.2	73.2	15.6	15.4	50.0	34.6
Planning Maze Test	33.61 ± 14.60	27.98 ± 11.88	0.65	14.2	77.7	8.1	23.9	71.8	4.3
Planning Maze Test total/time	0.54 ± 0.22	0.64 ± 0.28	0.61	7.5	90.2	2.3	4.3	93.5	2.2
TMT B-TMT A difference	50.02 ± 34.58	46.67 ± 40.26	0.67	6.0	78.2	15.8	0.0	56.0	44.0
Interference C. I. time	19.48 ± 5.01	20.70 ± 6.17	0.59	5.2	88.7	6.1	28.3	58.7	13.0
Interference C.I. total/time	1.83 ± 0.41	1.72 ± 0.49	0.77	4.3	90.1	5.6	8.7	82.6	8.7

Note. Mean and Standard deviation

HC Healthy controls, *rtt* Test-retest reliability, *AKT* Alters Konzentrations Test, *C.I.* Cerebral Insufficiency, *TMT A* Trail Making Test Version A, *TMT B* Trail Making Test Version B, *SWT* Semantische Wort-flüssigkeit, *PWT* Phonematische Wortflüssigkeit, *BNT* Boston Naming Test, *VSRT* Verbal selective reminding test

Appendix

Abstract

Cognitive functions play an important role in diagnosing dementia. Before clinical cognitive deteriorations appear, the progression of the disease can be predicted by neuropsychological tests. The aim of the study was to assess if the Neuropsychological Test Battery Vienna (NTBV) is a valid predictor for diagnosing dementia and if NTBV subtests can discriminate between the examined groups subjective cognitive decline (SCD), mild cognitive impairment (MCI), Alzheimer's disease (AD) and healthy control (HC). 358 patients, referred to a memory outpatient clinic because of cognitive deterioration, were analyzed. The same patients were surveyed in a follow-up assessment after a mean interval of 25.96 months to examine cognitive performance and disease progression. Psychometric criteria like internal consistency and test-retest reliability were determined and performances across diagnostic groups were compared. Receiver operator characteristics (ROC) curves and stepwise logistic regression were conducted to examine discrimination accuracy. Most diagnostic groups presented differences between NTBV subtests. ROC analysis showed that the area under curve (AUC) of subtests ranged from 0.20 to 0.91. Internal consistency revealed a high Cronbach's Alpha for the total sample. Pearson correlation for different diagnostic groups failed to reach a good test-retest reliability. In conclusion, NTBV can be regarded as an acceptable neuropsychological test for predicting dementia. The domain memory of NTBV achieved the best discrimination power for detecting dementia. Good psychometric criteria could be shown except for test-retest reliability. Therefore, diagnosis should be considered with caution.

Keywords: cognitive functions, neuropsychological testing, subjective cognitive decline, mild cognitive impairment, Alzheimer's disease

Zusammenfassung

Kognitive Funktionen spielen eine wichtige Rolle bei der Diagnose von Demenz. Bevor klinische kognitive Verschlechterungen sichtbar werden, kann die Krankheitsentwicklung mit neuropsychologischen Testverfahren vorhergesagt werden. Ziel der Studie war es zu untersuchen, ob die Neuropsychologische Testbatterie Vienna (NTBV) eine valide Vorhersagekraft besitzt, um Demenz zu diagnostizieren. Ebenso sollten die Klassifikationsfähigkeiten der NTBV-Untertests für die Diagnosegruppen bei subjektiver kognitiver Verschlechterung (SCD), leichter kognitiver Beeinträchtigung (MCI), Alzheimer Erkrankung (AD) und gesunden Kontrollen (HC) erfasst werden. Dafür wurden die Daten von 358 Patienten, die aufgrund kognitiver Beschwerden in eine Gedächtnisambulanz überwiesen wurden, analysiert. Die Patienten wurden zweimal in einem Intervall von durchschnittlich 25.96 Monaten getestet, um die kognitiven Fähigkeiten und die Krankheitsentwicklung zu erfassen. Psychometrische Gütekriterien wie die internale Konsistenz und die Test-Retest-Reliabilität wurden überprüft. Um die Klassifikationsgenauigkeit zu ermitteln, wurden Receiver-Operating-Characteristic-Analysen (ROC) und eine schrittweise logistische Regressionsanalyse berechnet. Unterschiede zeigten sich in fast allen Diagnostikgruppen hinsichtlich der NTBV-Untertest. Die Parameter „Area under curve“ (AUC) wiesen Werte zwischen 0.20 und 0.91 auf. Eine hohe interne Konsistenz für die gesamte Stichprobe konnte erzielt werden. Dagegen konnte keine zufriedenstellende Test-Retest-Reliabilität ermittelt werden. Zusammenfassend kann die NTBV als ein zufriedenstellendes Testverfahren gesehen werden, um Demenz vorherzusagen, wobei die Domäne „Gedächtnis“ die beste diagnostische Aussagekraft zeigte. Alle untersuchten Gütekriterien außer die Test-Retest-Reliabilität konnten als hinreichend bestätigt werden. Durch die fehlende akzeptable Test-Retest-Reliabilität sollten die Diagnosen mit Vorsicht betrachtet werden.

Keywords: kognitive Funktionen, Neuropsychologische Testverfahren, subjektive kognitive Verschlechterung, leichte kognitive Beeinträchtigung, Alzheimer Erkrankung

Instruction and evaluation of the NTBV

Introduction

First, the test conductor introduces the examination with the following sentence *"We will do an examination to see how well you can concentrate and how well your memory works. First, I have some general questions about your age and your education."*. The test persons are asked about their age, date of birth and schools they attended. Education is surveyed by counting all school and university years completed together. All apprenticeship years within the framework of training are not included in the calculation.

Semantic Verbal Fluency Test (SWT)

For the SWT, participants are instructed to name as many animals/ grocery items/ tools (in subsequent trials) as they can think of until the experimenter says "Stop". For each of the three categories, participants are given a time limit of 60 seconds. Word repetitions and wrong categorizations are subtracted from the total score. The number of remaining answers thus results in the test score of SWT. Mythical or fantastic creatures or animals (e.g. dragons) are not counted as correct in the category animals. The category grocery items can also include detergents and cleaning products. For the category tools, screw and nail are not included but scale and ruler are. If the participant asks whether a certain word can be listed, the test conductor should answer the question with "Yes" or "No". When there is still time left, the test person can be encouraged to continue with phrases such as *"We still have some time left, keep thinking"* or "*Can you think of any more words?"*.

Phonetic Verbal Fluency Test (PWT)

In the PWT, test persons are given an initial letter and are instructed to name as many words as possible starting with this letter. Two rules are established: firstly, proper names such as names for cities, countries, persons or companies do not apply. For example, for the letter S, participants may not use Salzburg, Sardinia, Sabine or Siemens. The second rule concerns word stem repetitions: For example, if the participant said the word "swim", they may not include words like swimming pool or swimming because this would be a word stem repetition. Apart from that, there are no further rules. The first trial starts with the letter B (*"as in Bertha"*), followed by letters F and L in subsequent trials. The test persons have 60 seconds for each trial, signaled by "*Start"* and "*Stop"* by the test conductor. The amount of correct answers results in the test score. False categorizations and word (stem) repetitions are subtracted from the total score. Wrong naming of categories can be corrected once, such as first names, city names or word stem repetitions. If there is time remaining, the test person can be encouraged to continue.

Verbal Selective Reminding Test (VSRT)

In the VSRT, the test conductor shows a list of grocery items to the test person. The test person is instructed to read the list out loud and try to remember the items. They are told to name the list afterwards, without regard to the order. After having gone through the whole list, the test conductor asks the test person to repeat the list and marks all named items with a check mark on the test sheet. The second trial continues with the following instruction *"I will show you the list again but only those words that you [the test person] have not said yet. Afterwards, you are asked to repeat all items again, including those you have already said in the first trial. Now, I will only show you the words that you did not say the first time"*. The test conductor then shows the test person the first item, ask him or her to read it out loud and remember it. After having gone through all the items that have not been named in the first trial, the test person is asked to repeat the whole list. The test conductor can remind the test person to also name the items that he or she already said in the first trial. This whole procedure is repeated for the third, fourth and fifth trial. A recall and recognition trial is conducted 20 minutes after the fifth trial (see below).

HAWIE-R: Digit-Symbol-Test

In the Digit-Symbol-Test, test persons are shown a sheet with boxes of numbers. The numbers from one to nine are coded with symbols. Their task is to put the corresponding symbol below every number. The test conductor shows the first two boxes of numbers as an example. The test person then can practice until the given line on the sheet. After the practice trial, the test person has 90 seconds to continue the task until the experimenter says stop. The number of correctly solved items results in the raw score and is calculated with an evaluation template. Misdrawn symbols are subtracted from the sum total of correctly solved items.

Alters Konzentrationstest (AKT)

This test starts with showing the test person a practice sheet which has semicircles on it. The test person is instructed to strike out all semicircles, that are facing up. Mistakenly struck out semicircles can be corrected by encircling the respective semicircle. The test conductor gives an example of the task and lets the test person practice it. Then, the test person is given a new testing sheet with the instructions to strike out all semicircles that are white on the left side, black on the right side and are facing up. One semicircle is struck out exemplary by the test conductor. The test person starts the task on signal of the test conductor. Mistakes made in the first line of the testing sheet can be pointed out to the test person. If the test person makes a cross when striking out, he is informed that striking out with one line is enough to complete the task. The time, the test person needs for the task is measured. The test score

(G) is calculated as follows: G = (35 + R – F), divided by time needed to finish the task. R refers to the number of correctly struck out symbols (maximum 20) and F refers to the number of wrongly struck out symbols. The test is canceled after 120 seconds.

Trail Making Test (TMT A)

The TMT A includes numbers which the test person is asked to connect in ascending order. The test conductor points out how to connect the numbers 1-3 (without drawing a line) and lets the test person practice for the numbers 1 – 8. The test person is then given a new sheet with the numbers 1 – 25. The test conductor shows the beginning and ending number, tells the test person to finish the task as fast as possible and then gives the start signal. If he or she makes a mistake, the experimenter points it out. The duration of correcting is included in the total time. If the test person crosses out the numbers, he or she is instructed to connect the numbers consistently. The time needed in seconds builds the total score. The test is cancelled after 180 seconds.

Trail Making Test (TMT B)

The TMT B follows a similar procedure as TMT A, only this time the test also includes letters. The test person is instructed to connect numbers and letters in an ascending order, like 1A, 2B, 3C, 4D. The test person practices the task before the test constructor gives her or him the task sheet. The time (in seconds) needed to finish this task represent the raw score of the TMT B. If the test person makes a mistake, the test constructor points it out while going back to the last right number or letter on the sheet. Here, it is allowed to give the test person support e.g. *"Which letter follows E?"*. The duration of correcting is included in the total time. Further, time of TMT A is subtracted from TMT B. The test is cancelled after 300 seconds.

5 Point Test

For the 5 Point Test the test person is shown a sheet with boxes each containing five dots. The task is to draw as many different dot patterns as possible in three minutes. The experimenter explains to the test person: "*A pattern is made by connecting for example two dots* (show example by connecting two dots on the left side of a box)*, which would represent a simple figure. It is also possible to draw more complex patterns by connecting all five dots* (show example). *It is important not to draw the same pattern twice and only to use straight lines".* The test starts when the test person has understood that task and with a starting signal by the test conductor. The test person is allowed to draw the two examples. The number of correct patterns/figures (without including repetitions or mistakes) is evaluated as well as the number of repetitions.

Planning Maze Test (NAI)

Here, the test person is instructed to navigate with a pen from the inside (marked with a star) of a drawn labyrinth to the outside (marked by another star). They are instructed to avoid blind alley and to always keep the pencil on the sheet. The time needed to get from the starting to the endpoint is evaluated as well as the number of errors. Errors are assessed with the help of an evaluation template that include the values of 1 and 2 and are summed up to a total error value (maximum of 16). The total score is calculated as follows: (16 – errors)/time needed. The task is cancelled after 120 seconds.

C.I. Symbol

The test person is shown a sheet with symbols (squares, stars and flowers). He or she is instructed to loudly count and point out all squares line by line as fast as they can. Test score is constructed by the time needed for the task in seconds. For every mistake made, one second is added. There are 44 squares in total. The test is terminated after 60 seconds.

C.I. Interference

The test person is shown a sheet with the letters A and B. Their task is to always read the opposite letter out loud, i.e. saying A for B and vice versa. After a learning trial, test persons are instructed to go through the whole sheet as fast as possible. If one letter is read wrong, the test person will be corrected, and the number of errors (E) is noted. The time needed (T) is assessed. The total score is calculated as follows: (34 – E) / T. The test is terminated after 60 seconds.

Stroop colour-word time colours.

The test person is shown a sheet with colours and is asked to name the colours as fast as possible, line by line. The time needed is assessed in seconds and noted as the total score. The test is terminated after 60 seconds.

Stroop colour-word time words.

The test person is now shown a sheet containing words written in different colours. He or she is instructed to name the colour, not the word. After the test conductor has demonstrated an example and the test person practiced an example trial, the test person is instructed to go through the whole sheet as fast as possible. The test conductor corrects the test person if he or she makes a mistake. The time needed for correction is included in the total time (assessed in seconds). Time needed (T) and number of errors (E) is noted. The total score / time is derived as follows: total (36 – E) / T. The test is terminated after 120 seconds.

Delayed recall and recognition (VSRT)

Approximately 20 minutes after the fifth trial of the VSTR, the instructions for delayed recall are given. The test person is asked to repeat as many items from the grocery list. The order of words is not essential. In case the test person does not remember the list, the item cards can be shown covered up. For the following recognition trial, the experimenter instructs the test person with *"I will now read you a grocery list and after every word, you shall say YES, if the respective word was on the grocery lists, and NO if it was not"*. No points are subtracted from the test score for repetitions of items. In case the test person already named all items on the list in a given trial, the subsequent trials can be skipped. The score for VSRT immediate recall is derived from the number of correct answers in the first trial. The score for learning performance (VSRT total recall) is calculated by adding the number of correct answers from the first to fifth trial. The score for delayed recall results from the number of correct answers. Intrusions are items that the test person named but were not on the list. The score for VSRT recognition is calculated by subtracting "false positives" from "recognized items". The sum of items that have been wrongly remembered as being on the lists is divided by 2 and results in the raw score "false positive".

Boston Naming Test (BNT)

For this test, the test conductor shows the participant images and asks at each image "*What is the name of this object?"* or *"What is this called?"*. The test person can turn over to the next image by himself. He or she has a maximum of 10 seconds to name the object on each image. If the test person is not able to name the image in the given time, the conductor encourages him or her and continues with the next item. If the participant makes a mistake and corrects him- or herself, the answer can be valued as correct. The total score is derived from the number of correctly named images. Mistakes are subtracted from the score. The test conductor can only provide support in case the test person gives an answer that is too unspecific, such as saying "boat" for the item "canoe" or saying "burning mountain" for the item "volcano". In this case, the test conductor can help by asking e.g. "*Is there any other name for this?"*. Examples for evaluation:

Item	Example for correct answer
flower	rose, thistle
house	school building, hospital
harmonica	mouth organ
camel	dromedary