universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Screening for Visuoconstructive Functions in Patients with Cognitive Impairment using the VVT 3.0 Delayed Recall – A Retrospective Analysis

verfasst von / submitted by

Jungwirth Patricia Bsc

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

Master of Science (MSc)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Claus Lamm
Mitbetreut von / Co-Supervisor:	Assoc.Prof. PD. Dr. Johann Lehrner

Acknowledgements

I would like to thank Ass. -Prof. Priv.-Doz. Mag. Dr. Johann Lehrner for offering me the opportunity to write my master´s thesis in such an interesting research field. Also, I would like to thank Univ.- Prof. Dr. Claus Lamm for his constructive feedback during my writing process.

Content

1. Theoretical Background

According to the World health organization (WHO), dementia is a syndrome caused by progressive brain disease, and features disturbances in several higher cortical functions (WHO, 2012). It affects many different cognitive domains such as memory, thinking, orientation, comprehension, calculation, learning capacity, language and judgment (Dilling, Mombour & Schmidt, 2011) and features progressive and significant deterioration (Gauthier et al., 2006). This cognitive decline is accompanied by behavioral and psychological symptoms of dementia such as impairments in daily living skills and related activities of daily living and deterioration in emotional control, social behavior, and/or motivation (DSM-5, 2013). According to the WHO (2020), dementia is one of the primary causes of disability and dependency among older people and the seventh leading cause of death among all diseases.

Consistent results of several studies show that the risk of developing dementia increases with age - people before the age of 60 are less likely to develop dementia, whereas the prevalence and incidence increases significantly after the age of 60 (Wancata, Musalek, Alexandrowicz & Krautgartner, 2003). The majority of those affected are older than 80 years and female (World Alzheimer Report, 2014). According to Wancata et al (2011), in 2013 64.307 people in Austria suffered from dementia, in 2050 it will be 262.200 people (Austrian Dementia Report, 2014). Worldwide, more than 55 million people live with dementia and due to the world's aging population this number is expected to rise to 78 million in 2030 (WHO, 2020). As a consequence, the massive impact the syndrome has on patients, families and society continues to increase, which is why the WHO has called for dementia to be a public health priority (WHO, 2012).

Dementia can have many different causes, whereby the most common cause is Alzheimer's Disease (Fiedler, Wiltfang, Peters & Benninghoff, 2012; WHO, 2012), which may contribute to 60 - 80% of all cases of dementia (McKhann et al., 2011).

1.1. Alzheimer's Disease

Alzheimer's disease (AD) is an irreversible and progressive neurodegenerative disorder. It is characterized by an increasing deterioration in cognitive performance, with impairment in episodic memory typically being the first to decline. As the disease progresses, deficits in semantic memory, language, executive functioning and visuospatial abilities develop (Weintraub, Wicklund, & Salmon, 2012). These deficits are usually accompanied by a

decline in the ability to cope with the activities of daily living (Dubois et al., 2014) and with increasing behavioural problems and neuropsychiatric symptoms (DSM-5, APA, 2013)

Even though the cause of AD remains unclear, several risk factors have been identified: increasing age (Hatcher, 1999; Lindsay & Anderson, 2004), subjective cognitive decline (SCD; Luck et al., 2015; Jessen et al., 2014; Mitchell et al, 2014), mild cognitive impairment (MCI; Forlenza, Diniz, Stella, Teixeira & Gattaz, 2013; Jessen et al., 2014; Mitchell, 2009), previous head injury, vascular disorders (Povova et al., 2012) and many lifestyle-related factors such as low educational attainment, physical and mental inactivity, obesity, high alcohol intake and smoking (Hersi et al., 2017; Solomon et al., 2014). According to Norton et al. (2014) and De Bruijn et al. (2015) about a third of risk factors can potentially be modified through enhancing educational opportunities or targeting the reduce of vascular risk factors, therefore reducing the incidence of AD. However, due to AD's complex multifaceted causes, that are still mostly unknown (Boyle et al., 2013) other therapeutical strategies are needed. Currently approved medications for treatment do not offer a cure yet, but simply alleviate the symptoms (Francis, Ramírez, & Lai, 2010).

Although the precise mechanisms leading to AD remain undefined, there are some characteristic neuropathological alterations in the brain during AD, that lead to cognitive decline.

1.2. Neuropathological Alterations and Biomarker Diagnostic

The cognitive decline typically starts in the brain's medial temporal lobe areas, crucial for memory (such as the entorhinal cortex and hippocampus) and extends to the frontal, temporal, and parietal lobes. The motor and sensory areas of the cortex, along with subcortical regions, however, are relatively unaffected in the early stages (Braak & Braak, 1991). Research suggests that amyloid-beta (Aβ) and tau proteins play a key role in the neurodegenerative processes observed in the brains of affected individuals (Karran, Mercken, & De Strooper, 2011). The pathological onset of AD probably starts when the Amyloid Precursors Protein (APP) is cut into smaller pieces, leading to the accumulation of toxic Aβ (Kumar, Singh, & Ekavali, 2015). Aβ leads to neuronal cell death by clumping together and sticking to neurons (Wang, Dickson, & Malter, 2006). Moreover, the accumulation of Aβ plaques leads tau proteins to misfold and form neurofibrillary tangles (NFTs), which impair neuronal communication and contribute to neuronal cell death (Hardy & Selkoe, 2002). While tau pathology was once considered a consequence of Aβ accumulation, it is also

possible that tau and Aβ may operate through parallel mechanisms, enhancing each other's toxic effects, while leading to AD (Small & Duff, 2008).

The neuropathological alterations that take place during AD and lead to serious neuron damage most likely begin years or even decades before any emotional, physical or cognitive symptoms of AD occur (Bondi, Salmon, & Kaszniak, 2009). Biomarkers focusing on those neuropathological features of AD are increasingly used to determine the likelihood of AD or MCI due to AD (Livingston et al., 2020). However, evaluating the effectiveness of these AD biomarkers is complex. This complexity arises because the range of neuropathological changes in AD patients can also be found in cognitively healthy elderly individuals who died from other causes (Curtis et al., 2015) and AD patients often exhibit various pathological changes (Kovacs et al., 2013). Furthermore, positive biomarker results indicate an increased risk for future cognitive decline, but they don't guarantee that cognitive impairment will develop within an individual's lifetime (Dubois et al., 2016). This variability in pathobiological manifestations prevents any biomarker from having 100% accuracy. Nonetheless, core cerebrospinal fluid (CSF) biomarkers like Aβ42, total tau (t-tau), and phosphorylated tau (p-tau) have been found to have a high diagnostic accuracy (85 – 90%) for identifying prodromal Alzheimer's in the MCI stage (Shaw et al., 2009, Visser et al., 2009) and they have a high negative predictive value (NPV), since negative amyloid results can exclude Alzheimer's pathology in individuals with cognitive decline (Livingston et al., 2020). Therefore, CSF biomarkers are useful in memory clinics for making diagnostic decisions (Duits et al., 2015). However, the process of acquiring biomarkers involves considerable expenses and invasiveness, and it necessitates the availability of imaging technologies or cerebrospinal fluid analysis (Albert et al., 2011). Furthermore, they do not assess cognition or identify cognitive dysfunction or symptoms, which the Alzheimer's Association emphasizes to do (Dubois et al., 2016).

Alternative to the biological approach, neuropsychological assessment has it's advantageous due to its practicality, versatility and lower costs (Jessen et al., 2014b). In this context methods and criteria for the objective detection of the earliest manifestation of cognitive decline are evaluated (Jessen et al., 2014c). The early detection of AD is crucial for slowing down the progression to dementia, since interventions are the most successful the earlier a cognitive decline is diagnosed and treatment started (Jessen et al., 2014b; Livingston et al., 2020). Instruments used for early detection of AD should be sufficiently sensitive to identify changes as the disease progresses.

1.3. Prodromal Stages

Jessen et al. (2010) suggested that dementia progresses in several stages: Subjective cognitive decline (SCD) is the early stage. SCD progresses to mild cognitive impairment (MCI) and eventually to dementia.

SCD as the slightest phase of cognitive decline, is characterized by patients noticing a subjective deterioration in different domains of cognitive functioning in relation to previous levels of cognitive functioning, while performing normal on standardized neuropsychological assessments adjusted for age, sex and education. SCD is a pre-clinical phase in which pathological changes leading to dementia are thought to be occurring but are not yet severe enough to produce objective cognitive impairment (Jessen, 2014).
Due to SCDs heterogenous nature (Karr et al., 2018) not all patients may progress to the stage of MCI or to dementia but remain stable or return to normal cognitive functioning. In a meta-analysis Mitchell et al. (2014) determined annual conversion rates of 6.67% from SCD to MCI and of 2.33% from SCD to dementia. Although the majority does not develop dementia (Jessen et al., 2020), there is a twofold higher risk for SCD patient than for controls without SCD (Mitchell et al., 2014; Jessen et. al; 2010&2014; Van Harten et al., 2018), making SCD a potential indicator for first manifestations of cognitive decline. Research on biomarkers (Jessen et al., 2018; Wolfsgruber et al., 2017) has shown that between 7% and 40% of people seeking medical attention for SCD are identified as being on the AD continuum, as evidenced by the presence of AD-related biomarker anomalies. Jessen et al. (2020) identified several risk factors that increase the likelihood of pre-clinical AD: experiencing subjective memory deterioration regardless of performance in other cognitive areas; beginning of SCD within the last five years; the emergence of SCD at or beyond the age of 60; concerns linked to SCD; ongoing SCD; consulting healthcare professionals and having cognitive decline verified by an external observer.

When patients transition into the MCI stage, general cognitive function is maintained (Albert et al., 2011), but an objective decline in neuropsychological performance, which is greater than expected for the affected person's age and education can be detected reliably (Gauthier et al., 2006; Nestor et al., 2004; Jessen et al., 2010). MCI is a diagnostic entity and according to the DSM-5 (APA, 2013) it is characterized by the following diagnostic criteria: "A. Evidence of modest cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual–motor, or social cognition) based on: 1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a mild decline in cognitive

function; and 2. A modest impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment. B. The cognitive deficits do not interfere with capacity for independence in everyday activities (that is, complex instrumental activities of daily living such as paying bills or managing medications are preserved, but greater effort, compensatory strategies, or accommodation may be required). C. The cognitive deficits do not occur exclusively in the context of a delirium. D. The cognitive deficits are not better explained by another mental disorder (for example, major depressive disorder or schizophrenia)." (APA, 2013, p. 605). Since cognitive decline may manifest in different cognitive domains (Jessen et al., 2014), four types of MCI can be distinguished (Petersen, 2004): Amnestic MCI (aMCI) which primarily affects the memory and learning and non-amnestic MCI (naMCI) which impairs other thinking skills, such as visuoconstruction, executive functions and language. Additionally, the diagnosis is categorized into MCI – single domain (diagnosed when only one cognitive domain is impaired) and MCI – multiple domain (diagnosed when more than one cognitive domain is impaired). According to a study conducted by Petersen et al. (2010) on a group of 1.969 elderly individuals aged 70 to 89 who were not diagnosed with dementia, 16% were found having any type of MCI (aMCI: 11.1%, naMCI: 4.9%). In another study he found that the prevalence of MCI increases with age: for ages 60 - 64 the prevalence of having MCI is 6.7%, for ages 65 - 69 it increases to 8.4%, for ages 70 - 74 to 10.1%, for ages 75 - 79 to 14.8% and for ages 80 - 84 to 25.2% (Petersen et al., 2018) However - like SCD - MCI outcomes are heterogenous and not every individual will proceed to dementia (Sachdev et al., 2014). Nonetheless MCI remains a risk factor for developing dementia, since there is some evidence that 14.9% of MCI patients older than 65 years have a two-year conversion rate to dementia, making the risk to develop dementia due to AD three times higher than for HC (Aisen et al., 2015). AMCI is more strongly related to conversion to AD than naMCI (single or multiple domain), whereas naMCI is more likely to proceed to other forms of dementia (Jessen et al., 2014). In a study by Fischer et al. (2007) it was observed that the rate of progression to AD over 30 months was significantly greater in the aMCI group (48.7%) compared to the naMCI group (28.8%). Similarly, Lehrner et al. (2016) identified a higher rate of conversion to AD from aMCI compared to naMCI in a clinical cohort, noting that among seven MCI patients who progressed to AD, five were from the aMCI group. The study indicated that individuals with aMCI were twice as likely to develop AD, as evidenced by an Odds Ratio of 2.0.

Eventually, cognitive impairment progresses to dementia. The DMS-5 (APA, 2013) renamed the term “dementia” into “major neurocognitive disorder” (major NCD). Major NCD is characterized by substantial cognitive deterioration severe enough to interfere with social or occupational functioning and consequently, a loss of independence. Additionally, behavior and mood might also be affected. Six distinct cognitive domains are named, that might be affected: the amnestic domain memory and learning and the non-amnestic domains complex attention, executive function, language, perceptual motor functioning and social cognition. To diagnose AD, impairments in at least two cognitive domains are necessary, with memory and learning being one of the affected domains, since they are usually the earliest and most significantly impacted cognitive domains.

Since the WHO identified early diagnosis as a key area of concern (WHO, 2012), reliable and valid instruments for early detection of SCD, MCI and dementia are required. For the diagnosis and early detection of dementia the DSM-5 (APA, 2013) and the National Institute on Aging and Alzheimer’s Association (NIA-AA; Hyman et al., 2012) mention visuoconstructive functions as an important cognitive ability to focus on. In the DSM-5 they belong as a subcategory to the non-amnestic domain of perceptual motor functioning.

1.4. Visuoconstructive Functions

Visuoconstructional functions involve understanding visuospatial relationships, planning and performing executive function skills and the coordination of fine motor skills with spatial abilities (Samrah et al., 2016). Deficits in visuo-constructional functioning are a frequent symptom related to dementia of the Alzheimer’s type (Freeman et al., 2000; Guérin et al., 2002) and have been found to decline early in the disease’s progression (Malloy, Belanger, Hall, Aloia, & Salloway, 2003; Possin et al., 2010). Many studies have found that assessing visuoconstructive functions serve as a valid means in distinguishing between subjects suffering from neurocognitive disorders and healthy controls (Lehrner et al., 2015; Valencia & Lehrner, 2018; Knechtl & Lehrner, 2023; Jacova, Kertesz, Blair, Fisk, & Feldman, 2007; Malloy, Belanger, Hall, Aloia, & Salloway, 2003). In addition, visuospatial abilities – as part of the broader spectrum of visuoconstructional functions – have been identified as predictors of fast cognitive decline (Buccione et al., 2007) and tend to decline more quickly than overall cognitive abilities in individuals developing AD (Karr et al., 2018**;** Johnson et al., 2009). These findings make visuoconstructional functioning important for neuropsychological screening instruments aimed at detecting incipient dementia.

However, a study conducted by Hansen et al. (2022) found that patients with cognitive impairment that are associated with cerebrospinal fluid (CSF)-based Alzheimer pathology demonstrated a preserved capacity in visuoconstructive tasks when verified with anti-neural autoantibodies (AP Aab+), in contrast to patients without autoantibodies (Aab–).This preservation suggests a potential moderating effect of autoantibodies on the typical decline in visuoconstructive functions observed in AD, offering a glimpse into the biological complexities underlying visuoconstructive functions. Such insights underline the nuanced and complex nature of AD and reinforce the importance of visuoconstructional assessment in capturing the heterogeneity of cognitive decline patterns.

According to the DSM-5 (APA, 2013) visuoconstructional functioning can be assessed in the reproduction of geometrical figures through drawing or copying that all demand hand-eye coordination. For instance, the clock drawing test (CDT; Shon et al., 2013) - which requires the subject to draw a clock with hands at a specific time without a time limit - can be used to assess visuoconstructional functions. For the CDT De Jager, Hogervorst, Combrenck, and Budge (2003) discovered that in comparison to control subjects, individuals with MCI were significantly impaired. Moreover Thomann et al. (2008) observed a significant gradual decline in clock drawing abilities across three groups: healthy controls, MCI patients and AD patients, with each group showing significant differences from the others. However, as a singular instrument the CDT has limited reliability in distinguishing between AD and its prodromal phases (Deuschl, Maier, et al., 2016; Breton, Casey, & Arnaoutoglou, 2019; Ehreke et al., 2010) and in predicting the conversion to AD (Amodeo et al., 2014) and is therefore only recommended in detection of dementia alongside the use of other tests.

The need for assessing visuo-constructive functions reliably, led to the development of the Vienna Visuo-constructional Test (VVT).

1.5. The Vienna Visuoconstructive Test (VVT)

The current version of the VVT (VVT 3.0) comprises an immediate/copy and delayed recall task. The immediate task requires the person to copy three figures, that each originate from another examination method, as accurately as possible: A clock, two overlapping pentagons and a three-dimensional cube. To assess delayed recall, after about 30 minutes the participant is asked to recall the same three figures from memory without any clues about the pattern and draw them as accurately as possible (delayed recall task).

The copy tasks of the VVT 3.0 predominantly concentrate on visuoconstructional abilities (Royall et al., 1998), including visuospatial abilities. Furthermore, sustained

attention, perceptual-motor ability and inhibition ability are required for the copying tasks, whereas free drawing tasks such as in the VVT 3.0 delayed recall require further executive control (Mendez et al., 1992; Royall et al., 1998) and memory retrieval (Caruso et al., 2020). In summary, the VVT 3.0 delayed recall focuses on assessing visuoconstruction, executive functioning, memory and learning.

For the use in diagnosis of AD and its prodromal phases the VVT 3.0 integrates three different visuo-constructional tests: The clock can be found in tests such as the Clock Drawing Test (Sunderland et al., 1989) and is one of the most commonly used figures when assessing visuo-constructional deficiencies with neuropsychological tests. Instead of freely drawing a clock, the person is required to copy it, due to the fact that copying provides a more accurate measure of visuo-constructive ability, without confounding effects of other abilities, such as planning or abstract thinking, which are required for a free drawing task (Pinto & Peters, 2009). The clock is set to "ten minutes after eleven" and its task and interpretation rely on that used in the Montreal Cognitive Assessment (MoCA), which evaluates different cognitive domains that are affected in dementia due to Alzheimer's (Nasreddine et al., 2005).

The two overlapping pentagons have been introduced into the Mini Mental State Examination (MMSE) and are also frequently used in the evaluation of neurocognitive disorders (Folstein et al., 1975). Pentagon copying tasks have been shown to be useful in distinguishing AD from other subtypes of dementia (Trojano & Gainotti, 2016), as well as in distinguishing AD from healthy controls (Nagaratnam et al., 2014).

The three-dimensional cube is used in the Alzheimer's Disease Assessment Scale - cognitive subscale (ADAS-COG) (Mohs et al., 1984). Like the MoCa, the ADAS-Cog is used to assess different cognitive domains that are impaired in AD. The total score of the ADAS-Cog correlates well with disease severity (Zec et al.,1992) and according to Salimi et al. (2019) patients suffering from AD have been reliably shown to perform worse in cube-copying tasks than HC.

Two scoring versions of the VVT exist: a full version and a screening version with fewer scoring items, which can therefore be scored faster. The main focus of this study will be on the VVT 3.0 screening version (Delayed recall). The maximum score for the screening is 10 points (3 for the clock, 3 for the overlapping pentagon and 4 for the three-dimensional cube). For the clock scoring includes correctly drawn contour, correctly placed digits and correctly placed pointers. For the two overlapping pentagons, scoring includes the drawing of two pentagons, their overlapping and that the overlapping part has four sides. Scoring for the

three-dimensional cube includes three-dimensionality of the figure, correctly oriented frontside, correctly drawn inner lines and that the opposite sides are parallel (within 10°). When participants draw the same pattern more than once, the best drawing will be graded. The Screening is scored in dichotomous yes/no (1/0), whereas zero indicates the worst performance. Scoring is done with an evaluation sheet and a transparent foil functioning as a stencil and takes approximately 30 seconds.

1.5.1. Former Research on the VVT 3.0 Screening and Delayed Recall

The VVT was developed by Lehrner et al. (2015) at the medical University of Vienna as a new valid and reliable instrument for assessing visuo-constructive functions.

Numrich et al. (2017) analyzed the ability of the VVT 3.0 Screening scores to differentiate between the patient groups HC, SCD, MCI and AD. Kruskal Wallis Analyses and additional pairwise post hoc comparisons revealed, that the scores were able to distinguish between AD patients and HC and all other diagnostic groups (SCD, MCI), and between HC and MCI. The scores were not able to differentiate between the groups HC-SCD and SCD-MCI. To determine if the VVT 3.0 screening scores can differentiate between patients with AD and those without AD, receiver operating characteristics (ROC) analyses were conducted, with AD as the positive condition. An Area Under the Curve (AUC) of .85 revealed that the screening scores demonstrate very good accuracy in distinguishing individuals with AD from those without AD. The Youden Index was used to determine an ideal cut-off score of 8.5. thus, patients that score lower than 8.5 in the VVT 3.0 Screening might suffer from AD. The cut-off revealed a sensitivity of .68 and a specificity of .93, thus 68% of individuals who had previously been diagnosed with AD by a clinician were accurately identified as having AD, while 93% of individuals were correctly identified not to have AD. Furthermore, the cut-off revealed PPV of .76, NPV of .89, LR+ of 9.71 and a LR- of .34.

Valencia and Lehrner (2018) also analyzed the ability of the VVT 3.0 Screening scores to differentiate between the diagnostic groups (HC, SCD, MCI and AD) in a cross-sectional analysis. Kruskal Wallis Analyses and Dunn's post hoc comparisons revealed that the scores of the screening version were able to differentiate between AD and all other diagnostic groups, but not between HC-SCD, HC-MCI and SCD-MCI. ROC analyses were conducted to assess the predictive validity of the VVT score for distinguishing diagnostic group membership, with AD as the positive condition. They revealed an AUC of .798 with an optimal cut-off at 8.5 points according to the Youden Index, which is equal to the cut-off that

Numrich obtained in her 2017 study. The cut-off delivered a sensitivity of .60 and a specificity of .86. Based on this cut-off PPV of .63, NPV of .84, LR+ of 4.18 and LR- of 0.47 were computed. As the screening version should primarily have a high sensitivity to achieve valid results in detecting patients with possible cognitive decline, Valencia and Lehrner (2018) suggested an alternative cut-off value at 9.5 points. This cut-off corresponds with a higher sensitivity of .78, a specificity of .64 and subsequent PPV of .53, NPV of .85, LR+ of 2.15 and LR- of .35. To further evaluate the predictive ability of the VVT 3.0 screening scores regarding group membership multinomial regression analyses were conducted. However, multinomial regression models revealed that group membership could only be predicted for 60.60% of all participants, namely for 92.70% of MCI patients and 44.30% of AD patients, but not for HC or SCD. Cohen's kappa as a measure of agreement between clinical diagnosis and diagnosis by VVT 3.0 Screening reached 0.25. Valencia and Lehrner (2018) suggest that this low agreement is probably caused by too similar scores in the HC, SCD and MCI groups. Visuoconstructive decline in those groups might be too subtle to be picked up by a screening measure. Novelty in Valencia's thesis were several longitudinal analyses conducted to predict disease progression. Of the original study sample, 110 patients were examined in a follow up. Two ROC analyses were conducted, to assess the ability of the VVT scores to predict disease progression from SCD to MCI and from MCI to AD. Both ROC curves revealed a low AUC of .525 for MCI and of .537 for AD, which indicates no discriminative ability (equivalent to random guessing), resulting in low sensitivity and specificity. However, these results should be interpreted with caution, due to the low progressions rates in Valencia's sample. Hence, to predict the VVTs predictive ability about disease progression in future longitudinal analyses, a bigger follow up sample with a longer time interval between examinations is recommended.

Another study about the VVT 3.0 screening was conducted by Tokaj and Lehrner (2022), where the VVT 3.0 Screening version was compared to the CDT (Sunderland) and the Mini Mental Status Examination (MMSE). The latter is widely used for a broad assessment of cognitive impairment and the severity of dementia by approaching multiple domains (Deuschl et al., 2016). The aim of this study was to evaluate the ability of the three tests in identifying group membership (HC, SCD, MCI and AD) and predicting disease progression. Kruskal-Wallis analyses revealed significant differences in performance across all four diagnostic groups for all three tests. To evaluate the ability of tests to differentiate between AD and all other participants ROC Curves were analyzed. For the VVT 3.0 Screening ROC curves revealed an AUC of .791 and therefore demonstrated a moderate ability to

differentiate between patients with AD and nonAD. The optimal cut-off score according to the Youden Index was set to 8.5, and a corresponding sensitivity of 62.1% and specificity of 83.1% (YI = 0.452, LR+ = 3.671, LR– = 0.456, PPV = 0.77, NPV = 0.76). Nagelkerke's was 0.34, and Cohen's kappa was .43 ($p < .001$). Adjusting the cut-off to 9.5 - as Valencia et and Lehrner (2018) proposed - increased sensitivity to 84.9%, while reducing specificity to 55.2%. The CDT yielded similar results, but both tests were outperformed by the MMSE, which achieved a high sensitivity (94.5%) and specificity (88.8%), making this multidomain approach a more accurate tool for diagnosis of AD. ROC curves aimed at differentiating between MCI and non-MCI, and SCD and non-SCD groups were also conducted. For MCI, the VVT 3.0 Screening reached an AUC of .6 with an optimal cut-off of 9.5, leading to a sensitivity of 49.2% and a specificity of 68.3% (YI = 0.175, LR+ = 1.55, LR– = 0.744, PPV = 82.2, NPV = 31.1, $p = 0.002$). For SCD, the VVT 3.0 Screening reached an AUC of .509, which means the diagnostic ability is barely better than chance. AUCs for MMSE and CDT were equally low, highlighting the challenges in early detection of SCD and MCI, which is partly due to their heterogenous nature. To assess the ability of the three tests in classifying participants into the four distinct diagnostic groups, multinomial logistic regression analyses were conducted. The results indicated that 60.6% of participants were accurately classified by the VVT 3.0 Screening. Cohen's kappa as a measure for agreement between clinical diagnosis and diagnosis by VVT 3.0 Screening reached .28, indicating a low level of agreement. This finding is consistent with Valencia & Lehrner's earlier study (2018), which reported a similar classification accuracy (60.7%) and a low Cohen's kappa of .25 and highlights the fact, that the test might not be able to accurately track the gradual decline in visuoconstructional skills. To improve detection, a solution could be to refine the scoring system to be more sensitive to smaller mistakes as signs of decline. In a longitudinal analysis of the study, ROC curves were generated to assess the ability of the tests to distinguish between stable participants and those who progressed and predict disease progressing within the group of MCI. The VVT 3.0 Screening revealed an AUC of .488, which is similar to that of the CDT (AUC of .497) and suggests that both visuo-constructional tests cannot be recommended to reliably predict disease progression. Solely the MMSE showed some validity in this task (AUC = .808 and a sensitivity of 78.6% and specificity of 71.0%) and might be helpful to get a hint of future progress. This study showed that the multi-domained MMSE surpassed the performance of tests focused solely on visuoconstruction, which raises the question of whether assessing only visuoconstruction is inadequate for accurately distinguishing between AD and other conditions. Valencia & Lehrner (2018) theorized that

the complex nature of AD cannot be fully captured through a single domain assessment. However, Tokaj concluded that choosing the VVT 3.0 Screening over the MMSE could be justified by several factors: The MMSE requires approximately 11 minutes for administration and necessitates adequate language skills from both the patient and the examiner for effective communication (Folstein et al. (1975). In contrast, the VVT 3.0 Screening is quicker to administer, taking just a few minutes, which can be significantly beneficial in a clinical environment. Additionally, the test is straightforward and does not rely on speech comprehension or the patient's native language, making them suitable for use in linguistically diverse settings.

Knechtl and Lehrner (2023) examined the full version of the VVT 3.0 delayed recall. The purpose of this study was to evaluate the differences in the VVT 3.0 delayed recall scores between HC and different diagnostic groups (SCD, MCI and AD) and to investigate the test's potential to differentiate between individuals with AD and without AD (HC, SCD, MCI). In this study, the DI-Quotient (delayed recall score/immediate copy score) was introduced as an additional VVT 3.0 score. The DI quotient controls for the different baseline levels of memory performance, since the delayed recall task also addresses memory and executive functioning in addition to viusoconstruction. Kruskal-Wallis tests revealed significant differences among groups for the VVT 3.0 delayed recall scores and the DI-Quotient. Dunn's post hoc analyses highlighted these differences as statistically significant between AD and all other diagnostic groups and HC, as well as between MCI and HC. However, comparisons between HC and SCD, and SCD and MCI did not reach statistical significance. To analyze the VVT's ability to differentiate between AD and nonAD ROC curves, ideal cut-offs and a logistic regression model were conducted. ROC analysis for the delayed recall score revealed an AUC of .890, identifying an optimal cut-off of 29.5 points with a sensitivity of 89.6% and specificity of 81%. For the DI-Quotient, the AUC was .870, with an optimal cut-off of .485 and a corresponding sensitivity of 88.7% and a specificity of 76.3%. Both AUCs can be classified as "excellent" according to the proposed standard of Hosmer and Lemeshow (2013). The binomial logistic regression model was able to classify 83.4% of AD patients correctly, demonstrating a sensitivity of 75.7% and specificity of 87%. The regression coefficient (B) of - .75 ($p < .001$) suggested that higher scores on the VVT 3.0 delayed recall were linked to lower chances of having AD. The model yielded a significant Cohen's Kappa of .619 ($p < .001$), indicating a substantial level of agreement (Landis & Koch, 1977) between the diagnoses according to the VVT 3.0 delayed recall scores and the diagnosis predicted by clinicians.

1.6. Conclusive Preview

In conclusion these four publications demonstrated the ability of the VVT 3.0 Screening and the VVT 3.0 Delayed Recall to discriminate between AD and other diagnostic groups (SCD, MCI) and HC. The purpose of the present study was to re-evaluate the findings of Knechtl and Lehrner (2023) and extending their research by focusing on the screening version of the VVT 3.0 delayed recall and incorporating two new diagnostic categories: “amnestic mild cognitive impairment” (aMCI) and “non-amnestic mild cognitive impairment” (naMCI).

The VVT 3.0 delayed recall not only evaluates visuoconstructive functions like the VVT 3.0 copy, but also addresses memory and executive function. Similarly, the DI-quotient reflects a range of cognitive functions, since it combines both scores. Thus, the VVT 3.0 delayed recall score and the DI-quotient could offer a more nuanced differentiation among various stages of the disease. Also, the Screening version has its usefulness, since it requires considerably less time than the full version, which can be beneficial in a clinical environment. Based on this the VVT 3.0 Screening (Delayed recall) might enhance the process of diagnostic staging and provide a useful instrument in the beginning of the diagnostic process, prior to the use of more complex, invasive and expensive tools.

2. Aims of the Study

The purpose of this study was to reevaluate the potential of the VVT 3.0 screening in evaluating visuoconstructional functions, with our main focus on the delayed recall. Additionally, we calculated the “DI-quotient” (delayed recall screening score/immediate copy screening score) to serve as an additional VVT 3.0 score.

Using the VVT 3.0 Screening (Delayed recall), performance in subjective cognitive decline (SCD), amnestic mild cognitive impairment (aMCI), non amnestic mild cognitive impairment (naMCI), Alzheimer’s disease (AD) and healthy control (HC) groups will be assessed to identify group differences in visuoconstructional functions and to determine whether VVT 3.0 screening scores (Delayed recall) can be used to distinguish between the groups HC, SCD, aMCI, naMCI (nonAD) versus AD. For this we defined adequate cut-offs for classification, which allowed us to calculate measures such as sensitivity, specificity, positive and negative predictive values (PPV/NPV) and positive and negative likelihood

ratios (LR+/LR-) to determine the effectiveness of the VVT 3.0 screening (delayed recall) score as a diagnostic tool for differentiating between groups.

A subgroup of these patients will be examined in a follow-up. The longitudinal design allowed for the determination of progression rates and the comparison of differences of the VVT 3.0 Screening scores between the first and second examination. Additionally, we compared scores of progressors and non-progressors.

Lastly, since all patients were subjected to an extensive neuropsychological evaluation prior to undergoing visuoconstructional testing, we investigated the correlation of VVT 3.0 screening scores and relevant sociodemographic and clinical variables: age, gender, years of education, global cognitive status examined with the Neuropsychological Test Battery Vienna (NTBV; Lehrner, Maly, Gleiß, Auff & Dal - Bianco, 2007), depressive symptoms examined with the Beck's Depression Inventory (BDI II; Hautzinger, Keller & Kühner, 2006) and premorbid IQ measured with the Wortschatz-Test (WST; Schmidt & Metzler, 1992).

2.1. Hypotheses

These aims serve as the foundation for the following questions and hypotheses:

Hypotheses 1: Do the VVT 3.0 screening scores (delayed recall) of the diagnostic groups HC, SCD, aMCI, naMCI and AD significantly differ?

Expectation: Based on multiple studies that demonstrated differences in visuoconstructional performance between different stages of AD, we expected to find significant differences in VVT 3.0 screening scores among all mentioned groups, except between HC – SCD and SCD – naMCI (Valencia & Lehrner, 2018; Tokaj & Lehrner, 2021; Knechtl & Lehrner, 2023). Since SCD by definition does not typically demonstrate declining performance on objective tests, scores between SCD and HC are presumed to be similar and since naMCI involves declines in cognitive areas other than memory, no significant differences in screening scores between SCD and naMCI groups are anticipated (Jessen et al., 2014).

Hypotheses 2: Are the VVT 3.0 screening scores (delayed recall) able to differentiate between HC, SCD, aMCI, naMCI versus AD (Cohen's Kappa > 0)?

Expectation: Based on the results of several studies, it is anticipated that AD patients will perform worse than all other groups (Valencia & Lehrner, 2018; Tokaj & Lehrner, 2021; Knechtl & Lehrner, 2023). Therefore, it is hypothesized that there is a cut-off that allows to

distinguish between HC, SCD, aMCI, naMCI and AD on the basis of the VVT 3.0 screening score (delayed recall), so that Cohen's Kappa shows at least a minimal degree of agreement (Cohen's kappa > 0).

Hypotheses 3: Do the differences of the VVT 3.0 screening scores (delayed recall) between the first and the second examination of the diagnostic groups HC, SCD, aMCI, naMCI and AD significantly differ?
Expectation: It is anticipated that the differences of the VVT 3.0 screening scores (delayed recall) between the first and the second examination would significantly differ between HC and all impaired groups (SCD, aMCI, naMCI and AD). While the performance of declined patients may improve, worsen or stay the same (Jessen et at, 2020), healthy controls are expected to maintain their performance.

Hypotheses 4: Do the VVT 3.0 screening scores (delayed recall) at the first examination significantly differ between progressors and non-progressors in the diagnostic groups SCD, aMCI and naMCI?
Expectation: It is anticipated that in the diagnostic groups SCD, aMCI and naMCI, progressors and non-progressors would have significantly different VVT 3.0 screening scores (delayed recall) during the first examination, with progressors having lower scores than non-progressors.

In addition to that, exploratory research questions were formulated asking whether there are associations between all three VVT 3.0 screening scores at E1 and several demographical and clinical variables included into this study (age, sex, years of education, objective cognitive performance, verbal IQ and depressive symptoms). Furthermore, participants who participated in our second examination were compared to those who didn't to explore possible dropout bias.

3. Methods

3.1. Data collection

Data of the present study has been collected in the research project called "The Vienna Conversion to Dementia Study" (VCD Study; EC-number: 174/2008) (Lehrner et al., 2015). The ultimate ambition of the VCD Study was investigating conversion rates from patients

with SCD and patients with MCI to dementia (Lehrner et al., 2016). The anonymized data of patients that had attended the Department of Neurology at the Medical University of Vienna will be analyzed retrospectively. The whole study population received information about the purpose of the examinations, the data processing including anonymization and signed a declaration of consent. In addition, participants have been informed that there is no risk of harm or distress to them and that they also have no ethical benefit from participating, but that there will be an increase in knowledge due to their participation that may contribute to better diagnosis, understanding, and possibly alleviation of suffering from dementia. Furthermore, Participants were informed that they could abort the testing session at any point without having to declare their reasons and without fearing any disadvantages as a result of this decision. The data are stored on a password-protected server of the Medical University of Vienna and were sent pseudonymised after approval of the study to me, Patricia Jungwirth.

3.2. Study Population

The study's sample consists of persons with cognitive dysfunction who either showed up to the neurological outpatient clinic of the Medical University of Vienna via self-referral due to subjective or objective memory declines or got referred by the Department of Neurology for further examination of their cognitive status. All patients who met the eligibility criteria and gave consent were consecutively included in the study. Eligibility criteria for diseased patients encompassed possible classification into one of the diagnostic groups (SCD, aMCI, naMCI, AD) and not meeting any of the exclusion criteria. Exclusion criteria should reduce possible confounders and were proven by anamnesis. They have been maintained from previous studies on the VVT and encompassed: (1) Neurological disorders such as cortical stroke, multiple sclerosis, epilepsy and other or traumatic brain injury in the past, which were determined by neuroradiological and clinical examinations. (2) Medical conditions possibly interfering with normal cognitive abilities including renal, respiratory, cardiac and hepatic disease. (3) Current major psychiatric disorder according to ICD-10 (Dilling et al., 2008) other than depressive symptoms. (4) Significant auditory, visual, language or motor deficits (Stephan, Brayne, Savva & Matthews, 2011, cited by Bodendorfer, 2015) (5) Other forms of dementia that are not related to AD (patients with neurological comorbidity) (6) Less than 50 years of age. The possible low achievement, despite appropriate cognitive function, is the cause for the fourth exclusion criterion. The exclusion criterions 1-5 were picked since they may each contribute to cognitive deterioration. To specify a minimum age, the sixth exclusion criterion was created.

3.3. Instruments

Mini Mental Status Examination (MMSE)

The MMSE (Folstein et al., 1975) was used as a screening tool to evaluate overall cognitive functioning (Deuschl et al., 2016) and was utilized for descriptive statistical analysis, with a maximum achievable score of 30 points. It is one of the tests used most frequently for dementia screening and provides a sensitivity of up to 81% and a specificity of up to 89% in detection of dementia (Tsoi et al., 2015). The assessment comprises 20 questions and brief tasks that evaluate a range of cognitive abilities (orientation, memory, the ability to follow verbal and written instructions). The MMSE takes about 5 minutes to complete and is scored by summing the points, with a higher score indicating better cognitive performance. According to Kessler et al. (2000) the MMSE has good interrater reliability (r = .83).

Vienna Visuoconstructional Test (VVT) 3.0 – Screening (Delayed Recall)

The VVT 3.0 screening as an instrument to assess visuoconstructional functions was already presented in detail in chapter “1.5. The Vienna Visuoconstructional Test (VVT)” and will be kept short: The test comprises two tasks: an immediate/copy task and a delayed task. In the copy task, participants were asked to copy three objects as accurately as possible: a clock set at 11:10, two overlapping pentagons and a three-dimensional cube. For the delayed recall task, after a delay of approximately 30 minutes the participant was asked to draw the three figures from memory as accurately as possible. No clue about the patterns was given.
Two scoring versions of the VVT 3.0 exist: One broad version and one screening version. Both the copy and the delayed recall task were scored with the screening version (10 scoring items resulting in 0-10 attainable points) and used to calculate the DI-quotient (delayed recall screening score/immediate copy screening score). The Study focuses on the VVT 3.0 screening (Delayed recall). Numrich et al. (2017) investigated the psychometric properties of the VVT 3.0 Screening and reported high internal consistency with a Cronbach's alpha of .84, indicating that the items in the test are measuring the same construct. Discriminant validity was tested through item-total correlation, and the results showed overall positive correlations ranging from 0.30 to 0.80 for each item with the screening score, implying that the test is measuring the intended construct. The test also demonstrated good interrater reliability (r = .84) and an average test-retest reliability of r = 0.54, which can be interpreted as appropriate for this test, since it indicates that the test results between the first and second examination

are not quite consistent, which is linked to the nature of AD due to its progression in cognitive decline. In general, the VVT 3.0 Screening demonstrated good psychometric properties and can be considered a reliable and valid measure of visuo-constructional functions.

Neuropsychological Test Battery Vienna (NTBV)

The NTBV (Lehrner et al., 2007) was specifically designed for dementia diagnosis in a clinical setting and is administered to assess the cognitive status of the patient (Pusswald et al., 2013). It combines various measuring tools to evaluate a wide spectrum of cognitive domains, namely memory, language, executive functions, psychomotor speed and attention (Lehrner et al., 2007). Several tests are used to evaluate these domains. Evaluation of the memory domain is conducted through the Verbal Selective Reminding Test (VRST; Lehrner et al., 2007), which includes subtests for immediate recall, total recall, delayed recall, and recognition of previously presented food items. The evaluate the domain language, the Phonematic and Semantic Verbal Fluency Test (Goodglass & Kaplan, 1983) and the modified version of the Boston Naming Test (BNT; Morris et al., 1989) are used. The assessment of executive functions is conducted using part A of the Trail Making Test (TMT; Reitan, 1979), the Stroop and Maze tasks from the Nürnberger Aging Inventory (NAI; Oswald & Fleischmann, 1997), the Interference subtest from the C.I. Test (Lehrl & Fischer, 1997) and the Five-Point Test (Regard et al., 1982). Assessment of the attention domain involves using the Alters-Konzentrationstest (AKT; Gatterer, 2008), part B of the Trail Making Test (TMT; Reitan, 1979), the score discrepancy between parts A and B of the TMT, the Digit-Symbol subtest from the WAIS-R (Tewes, 1994) and the symbol counting subtest of the Cerebral Insufficiency (C.I.) Test (Lehrl & Fischer, 1997). The disease progression often impairs all assessed domains sooner or later. Lehrner et al's research supported the idea that memory is the first cognitive domain to be damaged because it demonstrated the widest performance gap between AD patients and healthy controls. Attention, executive functions and language showed the second-largest decline. Despite this, the items of the NTBV have shown a significant discriminative power between the HC group and the AD group (Lehrner, 2007). Furthermore, the NTBV's sensitivity enables the diagnosis of prodromal stages of dementia, and it also has a good discriminating power for identifying AD (Lehrner et al., 2007). Lehrner et al. (2007) reported that the NTBV maintains high internal consistency, as evidenced by Cronbach's alpha values ranging from .87 to .89 among dementia patients. Moreover, the test-retest reliability of the NTBV, as indicated in studies by Hitzl (2015),

Lehrner et al. (2007), and Macher (2013), varies from .69 to .94, further establishing its reliability and validity as a diagnostic tool. Several tests from the NTBV, including the VSRT, the AKT, TMT A and TMT B and the digit-symbol-subtest of the German WAIS-R have demonstrated prognostic value in determining whether a patient who reports memory issues would progress to AD dementia within two years (Lehrner et al., 2016).
The NTBV offers standardized, validated and normative data adjusted for gender, age and education for each item. Completing the entire test battery takes around 45 to 60 minutes.

Wortschatztest (WST)

The WST by Schmidt & Metzler (1992) is a vocabulary test used to measure the verbal intelligence of participants. In this test, participants are presented with six words in a row, among which only one is a real word. Across 40 such sequences, the participant's task is to identify the real word. If Participants are unsure of the correct word, they are instructed not to guess. Each correct identification is scored with one point, leading to a maximum possible score of 40. From the raw scores, an intelligence quotient (IQ) can be derived. Schmidt & Metzler (1992) noted that the WST has excellent internal consistency, with a Cronbach's alpha of .92.

Beck Depression Inventory-II (BDI-II)

The assessment of depressive symptoms in this study was conducted using the German adaptation of the BDI-II (Hautzinger et al., 2009). The BDI-II consists of 21 Items regarding symptoms of depression and is a self-report questionnaire. Participants are required to select one statement out of four that most accurately reflects their feelings over the past two weeks. For scoring the points are summed up with a maximum possible score of 63, where a higher score suggests more severe depressive symptoms. Segal et al. (2008) have indicated good internal consistency for the BDI-II (Cronbach's alpha = .86) and have validated its convergent and discriminant validity.

3.4. Study Design and Testing Procedure

All patients received a standard neurological and neuropsychological evaluation conducted by neurologists, neuropsychologists and other study staff. Patients in the sample participated in an initial neuropsychological assessment (baseline) and a subgroup also participated in one follow-up examination at a later point in time. A certain amount of drop-out was anticipated because the minimum and maximum period between examinations was established at 12 - 48

months. The neuropsychological assessment included the MMSE as a screening tool to evaluate overall cognitive functioning (Deuschl et al., 2016), the VVT 3.0 and the NTBV to evaluate a wide spectrum of cognitive domains (Lehrner et al., 2007). MMSE scores are used to decide whether patients would be able to complete the entire NTBV, which takes approximately 45 minutes, or whether a shorter version, which takes approximately 20 minutes would be more appropriate. Participants achieving a MMSE score < 24 from 30 points were given the short version of the NTBV. For patients with moderate to severe dementia not scoring more than 14 points, the NTBV was fully dispensed with. If subjects are being presented with the long version of the NTBV due to their MMSE score (= > 24), they are additionally presented with the WST for an estimation of premorbid intelligence and the BDI-II for measuring Depression.

3.5. Diagnostic Group Classification

Patients are diagnosed and classified into diagnostic groups (SCD, aMCI, naMCI, AD) according to their performance in the NTBV during a neuropsychological evaluation performed in conjunction with a clinical interview. The diagnosis was set by neuropsychologists, neurologists and other study staff involved in the evaluation of the patient's cognitive status. The following diagnostic criteria were applied for classification into a diagnostic group: SCD was diagnosed according to the research criteria of the Subjective cognitive decline Initiative by Jessen et al. (2014), and MCI was diagnosed using Mayo clinic criteria (Petersen, 2004). Individuals classified with MCI were further categorized into aMCI and naMCI groups. A diagnosis of aMCI was given when the memory domain's z-score fell below -1.5 standard deviations (SD), regardless of deficits in other cognitive areas. Conversely, a naMCI diagnosis was made when the z-score in one or more non-memory cognitive domains dropped below -1.5 SD. Diagnosis of AD was based on the NINCDS- ADRDA criteria (McKhann et al., 2011) and the DSM-5 criteria of the American Psychological Association (DSM-5, 2013). Using flyers posted inside the Vienna general hospital, cognitively healthy subjects were recruited. They underwent a cognitive screening and a standardized clinical interview to determine their suitability. They must not exhibit or report any symptoms of cognitive deterioration and must not meet any exclusion criteria. The Mayo Clinic criteria were used to determine normal cognition in HC.

4. Statistical Analyses

Our statistical analyses were conducted using IBM SPSS Statistics for Mac, Version 29.0 and Microsoft Excel for additional calculations. The significance level (p) was set to 0.05 in all our analyses, therefore a $p \leq .05$ indicates a significant result. Analyses were performed on the entire study population and all available values. In case of any missing value, it was excluded from analyses on this specific variable. The sample size resulted from the availability of eligible and willing patients who came to the department of neurology within the time frame of data collection.

Statistical Methods on Sample Characteristics

Sample characteristics were described for the whole study population and each diagnostic subgroup at examination 1 (E1). Categorical variables such as gender and diagnostic group (HC, SCD, aMCI, naMCI, AD) were presented in absolute numbers and/or relative frequencies, whereas continuous variables such as age, education and test scores were presented descriptively by Mean and Standard Deviation (SD), and Median with Interquartile range (IQR)**.** Information on the follow-up sample at examination 2 (E2) was presented in the same way.

The focus of this study lies on the VVT 3.0 screening scores of the delayed recall and the DI-Quotient. However, in the analyses of our first and second Hypotheses the VVT 3.0 screening scores of the copy task are included, because they could provide further comparative insights and information on advantages of Delayed Recall scores and the DI-Quotient. In our Analyses the DI-Quotient Variable was computed by calculating the quotient of the delayed recall screening score and the immediate copy screening score (delayed recall screening score/immediate copy screening score)

Statistical Methods on the first Hypotheses

To test for normality in the distributions of all three VVT 3.0 screening scores of E1 (copy, delayed recall, DI-quotient) Kolmogorov-Smirnov tests were calculated. According to the significant results of Kolmogorov-Smirnov tests, all three VVT 3.0 screening scores did not follow a normal distribution: VVT 3.0 Screening (Copy): $D(639) = 0.287$, $p < .001$; VVT 3.0

Screening (Delayed Recall): $D(639) = 0.121$, $p < .001$; DI Quotient: $D(631) = 0.118$, $p < .001$. Therefore, non-parametric methods were chosen to be applied for correlation analyses and group comparison.

To test if the VVT 3.0 delayed recall screening scores of the diagnostic groups significantly differ, Kruskal-Wallis analyses were conducted. The ranks of all three VVT 3.0 screening scores were compared among the five diagnostic groups, whereas the latter represented the independent variable and the VVT 3.0 screenings scores were set as the dependent variable. To specify pairwise differences, these analyses were coupled with a Dunn's post hoc comparison test. For both, the Kruskal-Wallis analyses and Dunn's post hoc comparison the Bonferroni correction was applied, to make sure that statistically significant differences are reliable and not due to the increased risk of Type 1 errors, which is associated with multiple comparisons.

Statistical Methods on the second Hypotheses

To test whether VVT 3.0 screening scores are able to differentiate between non-AD and AD, receiver operating characteristic (ROC) curves for all three VVT 3.0 screening scores were computed, with AD defined as the positive condition.

First of all, we dichotomized our groups into a new binary variable so that HC, SCD, aMCI, and naMCI (nonAD) are coded as one category (=0) and AD as another (=1). To determine various cutoff points for the VVT 3.0 screening scores that best distinguish between non-AD and AD, ROC curves were computed, which display the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut-off points of a diagnostic test. The sensitivity refers to the ability of the test to correctly identify those with the disease, the specificity is the ability of the test to correctly identify those without the disease (Wirtz, 2019c). The ROC curve was interpreted by calculating the area under the curve (AUC), which provides a single measure of overall discriminative ability of the test (Linden, 2006). The AUC ranges from 0 to 1, whereby an AUC of 0.5 suggests no discriminative ability (equivalent to random guessing) and an AUC of 1.0 indicates perfect discriminative ability. The Youden Index (sensitivity + specificity – 1; Youden, 1950) was used to find the optimal balance between sensitivity and specificity, resulting in the best cut-off score for categorizing VVT 3.0 screening scores as indicative of AD or non-AD. This optimal cut-off was used to dichotomize the VVT 3.0 screening scores into a binary variable: Test values below the cut-off were counted as positive results, indicating the presence of AD (= 1), test values above the cut-off were counted as the absence of AD (= 0). To further investigate the effectiveness

of the VVT 3.0 delayed recall screening score as a diagnostic tool for differentiating between AD and non-AD, the chosen cut-off was applied to produce classification tables. These classification tables depict the number of patients that are correctly diagnosed as having AD (true positive [TP]) or not having AD (true negative [TN]), along with those incorrectly diagnosed as having AD (false positives [FP]) or not having AD (false negative [FN]). These values allowed for the calculation of positive and negative predictive values (PPV/NPV). The PPV (TP / [TP + FP]) refers to the probability that subjects with a positive test truly have the disease, whereas the NPV (TN / [TN + FN]) refers to the probability that subjects with a negative test truly don't have the disease (Wirtz, 2019a). Furthermore, positive and negative likelihood ratios (LR+/LR-) were calculated. LR+ (sensitivity / [1 – specificity]), provides an indication of how much the odds of the disease increase when a test is positive. A LR+ greater than 1 suggests that a positive test result is linked to the presence of the disease. LR- ([1 – sensitivity] / specificity), indicates the extent to which the odds of having the disease are reduced when a test result is negative. A LR- value less than 1 signifies that a negative test result is correlated with not having the disease (Gessner & Arndt, 2019a/2019b). To interpret the likelihood ratios the proposed guidelines of Jaeschke et al. (1994) were followed: (1) a LR+ > 10 and a LR- < 0.1 indicate strong diagnostic evidence (2) a LR+ > 5 - ≤ 10 and a LR- ≥ 0.1 - < 0.2 indicate high diagnostic evidence (3) a LR+ > 2 - ≤ 5 and a LR- ≥ 0.2 - < 0.5 indicates weak diagnostic evidence and (4) a LR+ > 1 - ≤ 2 and a LR- ≥ 0.5 - < 1 indicate scarce diagnostic evidence. Cohen's Kappa was determined to specify the degree of agreement between the diagnosis according to the optimal cut-off versus the diagnosis previously given by clinicians. For a more detailed look into the discriminating ability of the VVT 3.0 screening scores, namely the probability of having AD based on the VVT 3.0 delayed recall screening scores, binary logistic regression analyses were computed, with the VVT 3.0 delayed recall screening scores as independent variable and the diagnosis AD versus non-AD as binary dependent variable. Following this, again cross-tabulations were created to display the results of diagnosing by the aid of the model and Cohen's Kappa was determined to specify the degree of agreement between the diagnosis previously given by clinician and the diagnosis predicted by the logistic regression model.

Statistical Methods on the third Hypotheses

To test our hypotheses regarding the follow up (E2), first of all the time interval between E1 and E2 was investigated. A certain amount of drop-out was anticipated because the minimum and maximum period between examinations was established at 12 - 48 months. To analyze

whether the differences of the VVT 3.0 screening scores (Delayed recall) between the first and the second examination of the diagnostic groups significantly differ, we computed a new variable with those differences between the delayed recall screening scores (E2 minus E1) and conducted a Kruskal-Wallis analyses, whereas the diagnostic groups represented the independent variable and the differences between the delayed recall screening scores were set as the dependent variable.

Statistical Methods on the fourth Hypotheses

Lastly, we tested if the VVT 3.0 screening scores (Delayed recall) at E1 significantly differed between progressors and non-progressors in the diagnostic groups. Participants were classified as progressors or non-progressors based on changes in their diagnostic status between two examination points. If there was a worsening in their diagnostic category from the first to the second examination, they were categorized as progressors. Participants that remained in the same diagnostic group or improved were categorized as non-progressors.
A cross-tabulation was conducted to explore the relationship between progression status and diagnostic category. The VVT 3.0 screening scores (Delayed recall) of the diagnostic groups SCD, aMCI and naMCI at E1 were then compared between these two groups using the Mann-Whitney U test. Due to a large dropout rate, additional investigations concerning the screening scores of E1 and their association with disease progression, as well as on their predictive validity for disease deterioration through ROC analyses were dispensed with.

Additional Statistical Methods

In addition to that, exploratory research questions were formulated asking whether there are associations between all three VVT 3.0 screening scores at E1 and several demographical and clinical variables included into this study. Therefore, several Spearman's rank correlations were conducted (1) between all three VVT 3.0 screening scores and the MMSE, the WST, the BDI-II, years of education and age (As controls were not age-matched, correlations for age were separately conducted in HC and in the diseased groups) (2) between the VVT 3.0 screenings scores and all subtests of the NTBV (3) between the three VVT 3.0 screening scores, to measure and contrast the degree of overlap in the constructs they evaluate. For an estimation of the effect sizes of Spearman's correlation analyses, the conventions of Cohen (1988) were followed, which state that $r \geq .10$ indicates a small effect, $r \geq .30$ indicates a moderate effect and $r \geq .50$ indicates a strong effect.

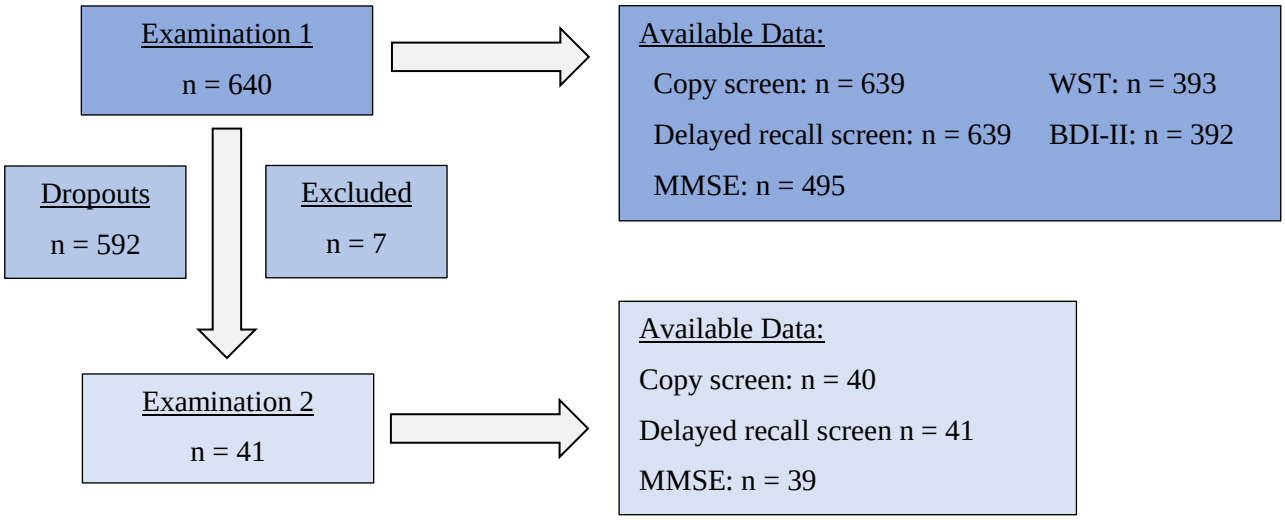
A Mann Whitney U test was used to explore correlations between gender and the VVT 3.0 screening scores, as well as to compare dropouts with those patients who participated in our second examination regarding the following domains: the VVT 3.0 delayed recall screening scores, the MMSE, years of education and age.

5. Results

5.1. Descriptive Statistics

Figure 1

Flow chart of patient recruitment at examination 1 and examination 2



The whole study population consisted of 640 patients and was divided into five diagnostic groups: 150 (23.4%) patients as HC, 49 (7.7%) patients with SCD, 114 (17.8 %) patients with aMCI, 166 (25.9%) patients with naMCI and 161 (25.2%) patients with AD. In total 296 (46.3%) men and 342 (53.4%) women participated in the first examination. Descriptive statistics for the whole sample and for the diagnostic subgroups can be found in Table 1 and in Table 2.

Table 1

Relevant sociodemographic and clinical variables at examination 1

	m/f	Age [a]	Education [a]	BDI-II	WST-IQ
HC	51/97	66.78 ± 10.11 66 (17)	13.78 ± 4.35 13 (6)	7.08 ± 7.22 5 (7)	108.15 ± 10.62 107 (17)
SCD	20/29	67.73 ± 10.28 67 (19)	13.20 ± 4.16 13 (9)	8.87 ± 7.07 7 (9)	112.22 ± 10.92 114 (18)
aMCI	64/50	70.25 ± 9.71 73 (16)	12.32 ± 3.91 11 (8)	13.19 ± 10.58 10 (16)	104.23 ± 13.33 104 (17)
naMCI	83/83	70.15 ± 9.31 72 (17)	12.38 ± 4.04 11.50 (7)	10.9 ± 8.24 9 (9)	106.40 ± 13.72 107 (21)
AD	78/83	74.80 ± 7.04 77 (10)	10.99 ± 3.45 11 (5)	9.13 ± 6.96 9 (9)	99.95 ± 16.14 99 (21)
N*	296/342	70.63 ± 9.58 73 (15)	12.38 ± 4.07 12 (7)	10.15 ± 8.57 8 (10)	106.43 ± 13.24 107 (19)

Note. M ± SD, Md (IQR). AD - Alzheimer's disease; aMCI - amnestic mild cognitive impairment; E1 – examination 1; HC - healthy controls; IQR – interquartile range; m/f – male/female; Md – Median; M (SD) - mean ± Standard deviation; N - absolute number; naMCI - non amnestic mild cognitive impairment; SCD - subjective cognitive decline. [a] in years. *For available Data for each test see Figure 1

Table 2

Cognitive screening scores at Examination 1

	MMSE[b]	VVT 3.0 Copy Screen	VVT 3.0 Delayed Recall Screen	DI-Quotient
HC	27.40 ± 0.55 27 (1)	9.47 ± 0.9 10 (1)	7.99 ± 1.2 9 (4)	0.85 ± 0.2 0.9 (0.6)
SCD	28.65 ± 1.76 29 (2)	9.51 ± 1.23 10 (1)	7.1 ± 2.93 8 (5)	0.74 ± 0.3 0.8 (0.5)
aMCI	26.53 ± 2.32 27 (3)	9.19 ± 1.21 10 (1)	4.31± 3.17 4 (4)	0.47 ± 0.33 0.44 (0.48)
naMCI	27.77 ± 2.15 28 (2)	9.25 ± 1.2 10 (1)	6.14 ± 3.09 6 (5)	0.65 ± 0.31 0.7 (0.46)
AD	21.26 ± 3.5 22 (4)	6.82 ± 2.92 8 (4)	1.24 ± 1.9 0 (2)	0.21 ± 0.36 0 (0.33)
N*	25.45 ± 3.99 26 (6)	8.70 ± 2.07 10 (2)	5.08 ± 3.63 6 (7)	0.57 ± 0.39 6 (7)

Note. M (±SD), Md (IQR). AD - Alzheimer's disease; aMCI - amnestic mild cognitive impairment; E1 – examination 1; HC - healthy controls; IQR – interquartile range; m/f – male/female; Md – Median; M (SD) - mean ± Standard deviation; MMSE – Mini-Mental State Examination; N - absolute number; naMCI - non amnestic mild cognitive impairment; SCD - subjective cognitive decline; VVT - Vienna Visuoconstructional Test. [b] in points. *For available Data for each test see Figure 1

The follow-up examination was set 12 - 48 months after E1. On average participants showed up to E2 18.6 months (SD 8.5) after E1. Of the original 640 cases, 48 took part in our follow-up examination. Since seven participants showed up to E2 prior to the 12-month minimum interval, they were excluded from further analyses. This was done to prevent the possibility of retained memory from the initial examination. The remaining 41 participants (6.25% of the total sample) were classified as follows: 3 (7.3%) healthy participants, 6 (14.6%) with SCD, 4 (9.8%) with aMCI, 15 (36.6%) with naMCI and 12 (29.3%) with AD. For one participant diagnostic group classification was missing. The follow-up sample consisted of 21 (52.5%) men and 19 (47.5%) women. More details on the sample characteristics and cognitive screening scores at E2 are displayed in Table 3. A flow-chart with the total numbers of participants at each test at both examinations can be found in Figure 1.

Table 3

Relevant sociodemographic and cognitive variables at examination 2

	m/f	Age [a]	Education [a]	MMSE[b]	Copy Screen	Delayed Recall Screen
HC	0/3	58.67 ± 8.62 51 (-)	16 ± 6.24 14 (-)	29 ± 1.41 29 (-)	9 ± 1.73 10 (-)	9.67 ± .58 10 (-)
SCD	4/2	70.67 ± 10.6 73.50 (14)	15.17 ± 4.96 16.50 (9)	29.33 ± .52 29 (1)	10 ± 0 10 (0)	9.67 ± 0.52 10 (1)
aMCI	2/2	74.75 ± 7.84 78.50 (12)	11.75 ± 5.68 27.50 (6)	26.75 ± 3.4 27.50 (6)	8.5 ± 1.29 8.50 (3)	4.5 ± 1.73 5 (3)
naMCI	8/7	66 ± 10.5 66 (19)	13.2 ± 4.96 11 (7)	28.53 ± 1.25 29 (2)	9.13 ± 1.36 10 (1)	7.13 ± 3.07 8 (5)
AD	7/5	70.67 ± 8.13 73.50 (14)	11.33 ± 4.03 10 (8)	22.17 ± 2.13 22 (4)	7.08 ± 3.18 9 (5)	2.36 ± 2.46 2 (4)
N	21/19	68.34 ± 9.74 71 (19)	12.88 ± 4.85 11 (8)	26.40 ± 3.54 28 (6)	8.59 ± 2.19 9 (3)	5.95 ± 3.7 6 (7)

Note. N = 40. AD - Alzheimer's disease; aMCI - amnestic mild cognitive impairment; E1 – examination 1; HC - healthy controls; IQR – interquartile range; m/f – male/female; Md – Median; M ± SD - mean ± Standard deviation; MMSE – Mini-Mental State Examination; N - absolute number; naMCI - non amnestic mild cognitive impairment; SCD - subjective cognitive decline; VVT - Vienna Visuoconstructional Test
[a] in years [b] in points
*Total sample: For available Data for each test see Figure 1

5.2. Results on the first Hypotheses

To test if the differences of the VVT 3.0 screening scores between the first and the second examination of the diagnostic groups HC, SCD, aMCI, naMCI and AD significantly differ, Kruskal-Wallis analyses were conducted coupled with a Dunn's post hoc analyses.

For the VVT 3.0 copy screening score the Kruskal-Wallis analyses depicted a significant difference between all diagnostic groups with $H(4) = 151.554$, $p < .0167$ and mean ranks of HC 379.86, SCD 409.47, aMCI 345.94, naMCI 358.49, and AD 179.11. Post hoc analyses using Dunn's procedure confirmed significant differences between AD and all other diagnostic groups ($p < .00167$). The remaining comparisons of the Dunn's post hoc analyses (HC - SCD, naMCI - SCD, naMCI - HC, aMCI - HC, aMCI - SCD, aMCI - naMCI) did not reveal significant results ($p > .00167$).

For the VVT 3.0 delayed recall screening score the Kruskal-Wallis analyses depicted a significant difference between groups with $H(4) = 302.043$, $p < .0167$ and mean ranks of HC 465.65, SCD 422.41, aMCI 279.54, naMCI 371.45 and AD 129.63. Dunn's post hoc comparison test found no significant differences between the groups SCD and HC ($p = 1.000$) and between SCD and naMCI ($p = .871$). Between all other groups Dunn's post hoc analyses showed significant pairwise differences (naMCI - HC, aMCI - HC, aMCI - SCD, aMCI - naMCI, AD - HC, AD - SCD, AD - aMCI, and AD - naMCI = $p < .00167$).

For the VVT 3.0 DI-quotient Kruskal Wallis analyses revealed significant group differences with $H(4)=249.685$, $p < .0167$) and mean ranks of HC 453.75, SCD 401.78, aMCI 268.64, naMCI 357.20 and AD 145.78. Dunn's post hoc analyses revealed the same results as for the VVT 3.0 delayed recall with no significant differences between the groups SCD - HC ($p = .813$) and between SCD - naMCI ($p = 1.000$) and significant differences between all other groups ($p < .00167$).

5.3. Results on the second hypotheses

To test if the VVT 3.0 screening scores are able to differentiate between HC, SCD, aMCI, naMCI versus AD (Cohen's Kappa > 0), ROC curves for all three VVT 3.0 screening scores were computed, and AD was defined as the positive condition.

For the VVT 3.0 screening (copy) ROC curves revealed an AUC of .795, 95% CI [.751; .839], $p < .001$. An ideal cut-off score of 8.5 points was determined, with a sensitivity of .847 and a specificity of .621. For the VVT 3.0 screening (delayed recall) the ROC curves presented an AUC of .898, 95% CI [.873; .923], $p < .001$. with an ideal cut-off of 2.5 points, a sensitivity of .856 and a specificity of .820. For the VVT 3.0 DI-Quotient the ROC curves revealed an AUC of .857, 95% CI [.822; .892], $p < .001$ with an ideal cut-off of .27 and a related sensitivity of .872 and a specificity of .724.

For the cut-off of 2.5 points of the VVT 3.0 delayed recall screening score positive (.653) and negative (.934) predictive values as well as positive (4.75) and negative (.210) likelihood

ratios were calculated. Since the VVT 3.0 screening (delayed recall) only allows full points, it is not possible that a test score equals the cut-off of 2.5, hence the score either over- or undershoots the threshold. Table 4 presents cross-tabulations for the diagnosis according to this optimal cut-off versus diagnosis given by clinicians. The degree of agreement between the diagnosis according to the cut-off and the diagnosis made by clinicians was defined by Cohen's Kappa as significant at .621 ($p < .001$).

Table 4

Crosstabulation: Diagnosis according to the VVT 3.0 delayed recall screening - cut-off < 2.5 versus diagnosis made by clinicians.

			Diagnosis made by clinicians		Total
			nonAD	AD	
Diagnosis according to the cut-off < 2.5	ADneg	n	409	29	438
		% within diagnosis according to the cut-off < 2.5	93.4% = NPV	6.6%	100.0%
		% within diagnosis made by clinicians	85.4% = specificity	18%	68.4%
		% of Total	63.9%	4.5%	68.4%
	ADpos	n	70	132	202
		% within diagnosis according to the cut-off < 2.5	34.7%	65.3% = PPV	100.0%
		% within diagnosis made by clinicians	14.6%	82% = sensitivity	31.6%
		% of Total	10.9%	20.6%	31.6%
Total		n	479	161	640
		% within diagnosis according to cut-off < 2.5	74.8%	25.2%	100.0%
		% within diagnosis made by clinicians	100.0%	100.0%	100.0%
		% of Total	74.8%	25.2%	100.0%

Note. AD - Alzheimer's disease; nonAD (HC - healthy controls, SCD - subjective cognitive decline, aMCI – amnestic mild cognitive impairment, naMCI – non amnestic mild cognitive impairment); n – absolute number; VVT - Vienna Visuoconstructional Test

In case of the VVT 3.0 screenings scores (delayed recall), the binomial logistic regression model was significant with a $X^2 = 429.937$ and a Nagelkerke $R^2 = .541$. In total 84.7% of cases were correctly classified by the model, with a sensitivity of 68.9% and a specificity of 90% (see Table 5). The positive predictive value was 0.698, the negative predictive value was .896.

Table 5

Classification table of the binomial logistic regression model

		Predicted* diagnosis made by clinicians		Percentage Correct
		nonAD	AD	
Observed diagnosis made by clinicians	nonAD	430	48	90.0
	AD	50	111	68.9
Overall Percentage				84.7

Note. * The cut value is .500
AD - Alzheimer's disease; nonAD (HC - healthy controls, SCD - subjective cognitive decline, aMCI – amnestic mild cognitive impairment, naMCI – non amnestic mild cognitive impairment)

In Table 6, the odds ratio for classification as AD based on the VVT 3.0 screening scores (delayed recall) is presented.

Table 6

Variables in the equation of the binomial logistic regression model

	B	SE	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
VVT 3.0 delayed recall screen	-.610	.051	142.947	1	.001	.543	.492	.600
Constant	.992	.165	36.209	1	.000	2.696	1.952	3.726

Note. CI – confidence interval, df - degree of freedom, Exp(B) – exponentiation of the coefficient B, SE – standard error, Sig. – p-value, VVT - Vienna Visuoconstructional Test

Cohens Kappa, as a measure of agreement between the diagnosis predicted by the logistic regression model and the diagnoses made by clinicians presented a significant Kappa of 0.591 ($p < .001$).

5.4. Results on the third hypotheses

In Table 7 the delayed recall screening scores of E1 and E2, as well as the differences between those scores (E2 – E1) are presented.

Table 7

VVT 3.0 delayed recall screening scores at E1 and E2

Groups	Scores at E1	Scores at E2	E2 – E1*
HC	9 (4)	10 (-)	-
SCD	8 (5)	10 (1)	0 (3)
aMCI	4 (4)	5 (3)	-.5 (5.5)
naMCI	6 (5)	8 (5)	0 (2)
AD	0 (2)	2 (4)	0 (3)

Note. Md (IQR). AD - Alzheimer's disease; aMCI - amnestic mild cognitive impairment; E1/2 – examination 1/2; HC - healthy controls; IQR – interquartile range; Md – Median; naMCI - non amnestic mild cognitive impairment; SCD - subjective cognitive decline; VVT - Vienna Visuoconstructional Test
*Diagnostic group refers to diagnosis at E1

To test if the differences of the VVT 3.0 delayed recall screening scores of the diagnostic groups between E1 and E2 significantly differ, a Kruskal Wallis Analyses was conducted, which did not reveal significant results: $H(4) = 435$, $p = .979$ and mean ranks for $HC_1 = 21$, $SCD_1 = 20.67$, $aMCI_1 = 22.08$, $naMCI_1 = 19.43$ and $AD_1 = 22.40$. Since no significant differences could be detected multiple pairwise group comparisons via Dunn's post hoc comparison were not performed.

5.5. Results on the fourth hypotheses

Lastly, we tested if the VVT 3.0 screening scores (Delayed recall) at E1 significantly differed between progressors (n = 12, 30%) and non-progressors (n = 28, 70%) in the diagnostic groups. Mann-Whitney U test did not prove significant differences between the VVT 3.0 delayed recall score of progressors and non-progressors in none of the diagnostic groups. However, analyses included only a small number of participants (n = 40).

5.6. Additional Statistical Methods

Spearman's rank correlation analyses were used to explore the associations between the three VVT 3.0 screening scores and relevant sociodemographic, clinical and cognitive variables (see Table 8). Spearman's rank correlations between all subtests of the NTBV and all three VVT 3.0 screening scores reveal that all pairwise comparisons were significant. In total Spearman's rho values range from - .55 to .656.

Table 8

Spearman's rank correlations of the VVT 3.0 screening scores at E1 with age, the MMSE, the WST, the BDI-II, years of education and the subtests of the NTBV

			VVT 3.0 copy	VVT 3.0 delayed recall	VVT 3.0 DI-quotient
Spearman's rho	Age*[a] (n = 490)	r	-.208**	-.368**	-.354**
	Age*[b] (n = 149)	r	-.216**	-.165*	-.110
	MMSE (n = 495)	r	.512**	.666**	.598**
	WST (n = 392)	r	.199**	.195**	.166**
	BDI-II (n = 392)	r	-.068	-.067	-.077
	Years of education	r	.253**	.277**	.238**
	NTBV: *Domain attention*				
	AKT time (n = 486)	r	- .471**	- .528**	- .457**
	AKT total/time (n = 484)	r	.48**	.545**	.475**
	Digit-symbol test (WAIS-R) (n = 363)	r	.306**	.457**	.4**
	TMT B (n = 351)	r	- .446**	- .55**	- .485**
	NTBV: *Domain psychomotor speed*				
	symbols counting (c.I.) (n = 477)	r	- .402**	- .408**	- .336**
	TMT A (n = 483)	r	- .494**	- .538**	- .461**
	NTBV: *Domain language*				
	Semantic fluency (SWT) (n= 370)	r	.297**	.522**	.463**
	Phonematic verbal fluency total (n = 363)	r	.254**	.309**	.279**
	NTBV: *Domain memory*				
	Verbal memory total recall (VSRT) (n = 486)	r	.295**	.495**	.458**
	Verbal memory immediate recall (VSRT) (n = 485)	r	.372**	.654**	.622**
	Verbal memory delayed recall (VSRT) (n = 484)	r	.344**	.656**	.624**
	Verbal memory recognition (VSRT) (n = 484)	r	.295**	.563**	.537**
	NTBV: *Domain executive functions*				
	Nonverbal fluency (5 point test) total (n = 352)	r	.394**	.505**	.444**
	Nonverbal fluency (5 point test) perservations (n = 350)	r	- .131*	- .208**	- .2**
	Stroop Test colors time (n = 349)	r	- .23**	- .308**	- .259**
	Stroop Test words time (n = 348)	r	- .347**	- .419**	- .357**
	Stroop Test total/time (n = 348)	r	.358**	.405**	.34**
	Stroop Test interference (III-I) (n = 346)	r	- .307**	- .366**	- .312**
	Planning maze test (NAI) time (n = 483)	r	- .492**	- .51**	- .425**
	Interference (TMTB-TMTA) (n = 350)	r	- .421**	- .5**	- .444**
	Interference time (c.I.) (n = 482)	r	- .512**	- .523**	- .441**
	Interference total/time (c.i.) (n = 483)	r	.529**	.531**	.440**

Note. AKT - Alters-Konzentrations-Test; C.I. - cerebral Insufficiency test; BNT - Boston Naming Test; N – absolute number; BDI-II - Beck Depression Inventory-II; E1 – Examination 1; MMSE - Mini-Mental State; NAI - Nürnberger Altersinventar; NTBV – Neuropsychological Test Battery Vienna; N – absolute number; PWT - phonematic verbal fluency; r – Spearman's rho/ correlation coefficient; SWT - semantic verbal fluency; Sig. – p-value/statistical significance (2-tailed); TMT A - Trail Making Test version A; TMT B - Trail Making Test version B; VSRT - Verbal Selective Reminding Test; VVT – Vienna Visuoconstructional Test, WST – Wortschatztest
*[a] Analysed population includes only diseased patients. *[b] Analysed population includes only HC.
* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Spearman's rank correlations between the three VVT 3.0 screening scores show significant associations between every VVT 3.0 screening score (see Table 9).

Table 9

Spearman's rank correlations between the three VVT 3.0 screening scores

Spearman's rho		VVT 3.0 copy screen	VVT 3.0 delayed recall screen
VVT 3.0 delayed recall screen	r	.507**	
	N	638	
VVT 3.0 DI-quotient	r	.324**	.955**
	N	631	631

Note. N – absolute number; r – Spearman's rho/ correlation coefficient; VVT – Vienna Visuoconstructional Test.
** Correlation is significant at the 0.01 level (2-tailed).

Table 10 lists the results of our Dropout Analysis and compares dropouts to follow-up participants regarding sociodemographic, cognitive and visuoconstructive characteristics.

Table 10

Descriptive and post hoc comparisons regarding relevant variables at examination 1 between Follow-ups and Dropouts

	U	Z	Sig.*	Follow-ups**	Dropouts**
Copy Screen	11746.000	- .091	.928	8.80 ± 1.74 9.5 (2)	8.69 ± 2.1 9 (2)
Delayed Recall Screen	10619.500	- 1.333	.182	5.73 ± 3.84 6 (8)	5.04 ± 3.61 5 (7)
DI-Quotient	9611.000	- 1.888	.059	.66 ± .39 .83 (.7)	.56 ± .39 .6 (.7)
MMSE	6610.500	-2.781	.005	27.20 ± 2.42 27.50 (5)	25.3 ± 4.06 26 (7)
Age	10470.500	- 1.472	.141	68.34 ± 9.75 71 (19)	70.56 ± 9.52 73 (15)
Education	11518.000	- .265	.791	12.88 ± 4.85	12.36 ± 4.01

				12 (7)	12 (8)

Note. IQR – interquartile range; M – Mean; Md - Median; MMSE - Mini-Mental State Examination; SD - Standard Deviation; U – Mann-Whitney U test statistic; VVT - Vienna Visuoconstructional Test; Z – standard score.
*2-tailed **M ± SD, Md (IQR)

We found no significant differences between Dropouts and Follow-up participants except for the MMSE scores: Follow-up participants scored significantly higher in the MMSE than dropouts.

6. Discussion

6.1. Summary and Interpretation of Findings

In summary, our statistical analysis provided insight into the ablities of the VVT 3.0 screening; on the one hand to evaluate visuoconstructional functions alongside memory and executive functions, on the other hand to reveal information about cognitive decline in different stages, namely HC, SCD, aMCI, naMCI and AD. We explored the differences between these stages and the effectiveness of the VVT 3.0 screening in identifying group membership. Furthermore, we carried out longitudinal analyses over a period of 12 - 48 months to observe changes in performance over time. Finally, we examined the relationships between the VVT 3.0 screening scores, various clinical and sociodemographic variables and changes in performance across time.
In the following, our results will be discussed, references to other studies will be created and limitations will be illustrated to further clarify the possible future application of the VVT 3.0 screening.

Descriptive Statistics of the sample

In our study sample, the majority were patients with AD (25.2%) and MCI (aMCI: 17.8%; naMCI: 25.9%), which was anticipated given that the department of Neurology is predominantly approached by individuals experiencing cognitive decline, who are either referred for further evaluation or seek help on their own initiative. SCD (7.7%) made up the smallest group in our sample which is not surprising either since it is challenging for patients that might suffer from SCD to distinguish between the natural cognitive decline associated with aging and the abnormal deficits indicative of a disorder. Given this context, actively recruiting more SCD patients could have potentially led to a more even distribution and

increased sample size of SCD participants in the study. Nonetheless, Jessen, Amariglio, Buckley, et al. (2020) specifically advised against proactively seeking out individuals who do not independently approach medical facilities for assessment. The gender distribution across the entire sample (Men: 46.3%; women: 53.4%) and within diagnostic groups was fairly balanced. Education levels decreased over the course of the diagnostic group progression, being the highest in the HC group (mean 13.78) and the lowest in the AD group (mean 10.99). This is consistent with the common assumption that a lower level of education may be a risk factor for cognitive impairment and may contribute to the progression of AD. In terms of age, we observed that the average age increases across the diagnostic categories, with individuals with AD (mean 74.80) being the oldest. This pattern points out the gradual nature of cognitive decline associated with AD and highlights age as a significant risk factor (Dubois et al., 2016). Moreover, the MMSE scores that are indicative of the extent of cognitive impairment, decline across the diagnostic groups, with the lowest scores in the AD group (mean 21.26). That finding aligns with our expectations based on the progressive nature of the disorder. However, in our sample, the HC group (mean 27.40) showed a lower MMSE score than the SCD group (mean 28.65). One reason for that might be that the Individuals in the SCD group are more focused when taking the test due to their concerns about their cognitive function, leading to higher performance. Another reason could be that the HC group potentially included individuals with subjective cognitive decline that lowered the average MMSE score, since distinguishing between SCD and normal cognitive functioning is challenging due to the variability in perception of cognitive decline and the lack of clear, objective thresholds for defining cognitive decline (Jessen et al., 2020).

VVT 3.0 Screening scores across diagnostic groups

When examining the medians of all three VVT 3.0 screening scores, we observed that the decline was more pronounced in the VVT 3.0 screening (delayed recall) compared to the VVT 3.0 screening (copy). Similarly, the DI-quotient appeared to primarily follow the distribution pattern of the delayed recall scores and showed minimal reliance on the copy scores. For the delayed recall we can observe a progressive decline in median scores as cognitive impairment increases from HC to AD. Since the delayed recall task in addition to visuoconstructional abilites also requires executive functioning (Mendez et al., 1992; Royall et al., 1998) and memory retrieval (Caruso et al., 2020) the lower scores observed in delayed recall can be attributed to memory and learning difficulties (Albert et al., 2011), which are often the first cognitive abilities to decline in AD (DSM-5, 2013). Regarding MCI, both

groups show signs of cognitive decline, however the impact on delayed recall is more severe for patients with aMCI compared to naMCI. This could be attributed to the core symptomatology of each subtype: While aMCI patients are primarily impaired in the memory domain, naMCI is characterized by impairments in other cognitive domains such as language, attention, or executive function (Petersen, 2004). Therefore, naMCI patients, whose memory may be less affected, perform better on this task. The median scores and IQR of the VVT 3.0 screening (copy) indicate minimal variation among HC, SCD, aMCI and naMCI groups. All those groups achieved a median of 10 with an IQR of 1. However, there are also outliers present in those groups. It's unclear why some individuals in the HC and SCD groups displayed poor visuoconstructive skills. Possible explanations could include a lack of motivation, insufficient attention, or limited drawing skills, but further investigation is needed to understand these anomalies. The AD group achieved a median of 8 with an IQR of 4, which suggests a decline in visuoconstructional ability and increased variability, typical of Alzheimer's progression. The consistent medians across the HC, SCD, aMCI, and naMCI groups could suggest that the VVT 3.0 screening (copy) may not be sensitive enough to distinguish between these stages of cognitive decline in terms of visuoconstructional abilities, or that these abilities are not as affected in the early stages of cognitive decline as they are in AD.

Score Differences between groups

Regarding our first hypotheses, the Kruskal Wallis analyses revealed significant results for all three VVT 3.0 screening scores. Therefore, we can accept the alternative hypothesis which states that the scores of the diagnostic groups HC, SCD, aMCI, naMCI and AD significantly differ. This finding is consistent with previous findings on the VVT 3.0 Screening (Valencia & Lehrner, 2018; Tokaj & Lehrner, 2021) and VVT 3.0 delayed recall (Knechtl & Lehrner, 2023). Dunn’s post hoc comparisons to specify pairwise differences between groups revealed that all our expectations were met: No significant differences were found between the groups SCD and HC and between SCD and naMCI, for both the delayed recall screening score and the DI-Quotient. Since SCD by definition does not typically demonstrate declining performance on objective tests (Jessen et al, 2014b), VVT 3.0 screening scores in SCD were not distinct enough from HC. This observation reinforces the importance of appropriate diagnosing criteria for SCD, as it can help spot those who might be on the path to more serious cognitive issues, even when their test results still look normal. However, according to Jessen (2014) the accuracy of objective cognitive tests in detecting cognitive decline

improves as the severity increases; thus, at early stages when the decline is mild, distinguishing between normal and impaired cognition is less reliable. Therefore, distinguishing between HC and SCD remains challenging (Jessen et al., 2020). The absence of significant differences in the scores between SCD and naMCI was anticipated since naMCI involves declines in cognitive areas other than memory, such as visuoconstruction, executive functions and language. Those deficits might be too subtle to be captured by the screening of the VVT 3.0., which might be because naMCI can occur in multiple domains and can vary widely across individuals; also, impairments in those cognitive domains tend to occur later in the disease progression (Jessen et al., 2014) further complicating the detection of significant differences using a single neuropsychological screening measure. In general, the VVT 3.0 Screening (delayed recall) might be not sensitive enough to detect the more subtle cognitive changes that differentiate HC, SCD and naMCI.

Ability to differentiate between groups

To continue with our second hypothesis, the differentiation between AD and nonAD by the screnning scores, ROC curve analyses were computed. For the delayed recall they revealed an AUC of .898 and an ideal cut-off of 2.5 points. According to sensitivity and specificity values 85.6% of patients with AD are also classified as such, and 82% of all nonAD patients are identified correctly when the cut-off is used for diagnostic group assignment. Regarding the DI-quotient, the AUC was a little bit lower at .857, with an ideal cut-off at .27 and a sensitivity of 87.2% and a specificity of 72.4%. The VVT 3.0 delayed recall screening score and the DI-quotient both outperformed the differentiating ability of the copy screening score, which revealed the lowest AUC of .795 and lower values of sensitivity (84.7%) and specificity (62.1%). These findings suggest that the delayed recall screening has a higher diagnostic accuracy compared to the DI-quotient and the copy screening, with the AUC for the delayed recall and DI-quotient being classified as "excellent" and the AUC for the copy task as "acceptable," according to the standard proposed by Hosmer and Lemeshow (Mandrekar, 2010). The increased discriminatory ability of the delayed recall scores compared to the copy scores, could be due to the inclusion of memory and executive functions in the delayed recall task, since they worsen in the disease progression of AD, with memory typically being the first domain to be impaired (DSM-5, 2013; Lehrner et al., 2005) In addition to that crosstabulations revealed, that 65.3% of the patients who scored below 2.5 points were indeed diagnosed with AD (PPV) and 93.4% of those who scored above this threshold were diagnosed as not having AD (NPV). Unlike the PPV and NPV, the likelihood

ratios are not influence by the distribution of AD versus nonAD in our study sample and allow the prediction of the chance to have AD based on the delayed recall screening scores (Deeks & Altman, 2004): The likelihood ratio of scoring below 2.5 points was 4.75, suggesting that AD patients are 4.75 times more likely to score below this cutoff compared to nonAD patients (LR+). The likelihood of AD patients receiving a negative test result is .210 times more likely than for patients identified as HC, SCD, aMCI or naMCI (LR-). While these results are promising, caution is advised in their interpretation. The cut-off was determined post hoc using the Youden Index, which strives for optimal sensitivity and specificity. This could lead to an overestimation of test performance (Ewald, 2006; Leeflang et al., 2008). The chosen cut-off in our sample might not achieve the same level of accuracy in different populations or be the optimal cut-off. Thus, alternative cut-offs should be considered based on the specific objectives of its use. Adjusting the cut-off is necessary if the diagnostic goal shifts, such as aiming for higher sensitivity and PPV, which may require accepting lower specificity and NPV (Cohen et al., 2016). The strength of interrater reliability between the diagnosis according to the cut-off of 2.5 points and the diagnosis AD versus nonAD made by clinicians was defined by Cohen's Kappa as significant at .621 ($p < .001$). Following the interpretation of Landis and Koch (1977), this value can be graded as "substantial agreement".

Next to ROC analyses and ideal cut-offs, we calculated binomial logistic regression models, to further evaluate the potential of the VVT 3.0 screening (Delayed recall) to discriminate between AD and nonAD. The binomial logistic regression model delivered satisfactory results. The Nagelkerke R^2 indicated that the model explained 54.1% of the variance of the diagnosis nonAD versus AD by the VVT 3.0 delayed recall score. Altogether the model predicted 84.7% of the group membership correctly, 68.9% of the AD patients and 90% of the nonAD patients. 69.8% of the patients that the model diagnosed as AD patients actually had AD (PPV), 89.6% of the patients that the model classified as nonAD patients, indeed, did not have AD. A beta coefficient of - .610 ($p < .001$) suggests that higher scores on the VVT 3.0 screening (delayed recall) correlate with lower chances of having AD. When inserting this value into the formula of the natural logarithm with the Euler's number as a base, it results in .543. This means that for each additional point on the VVT 3.0 delayed recall score, the odds of having AD are multiplied by .543, or to interpret this odds ratio as a percentage decrease: for every point increase in the VVT 3.0 delayed recall score, the odds of having AD decrease by 45.7%. Moreover, the Cohen's Kappa as a measure of agreement between diagnosis made by the logistic regression model's and the diagnosis made by clinicians,

reached “moderate agreement” at .591 (with $p < .001$) (Landis and Koch, 1977). Concluding our second hypotheses, we would again accept the alternative hypothesis (H1), which states that VVT 3.0 delayed recall screening scores can distinguish between HC, SCD, aMCI, naMCI and AD. Our findings of the VVT 3.0 screening for the delayed recall task resemble those of the VVT 3.0 full version for the delayed recall conducted by Knechtl and Lehrner in their 2023 study. In conclusion, the VVT 3.0 screening (delayed recall) appears to be a very effective tool for screening individuals for AD. It has demonstrated a strong ability to correctly classify individuals as having AD or not, with a high degree of accuracy. The model's predictive values and the significant beta coefficient all support the conclusion that the VVT 3.0 screening (delayed recall) could be a valuable component in the diagnostic process for AD, helping to differentiate between those with AD and those without.
However, it's important to note that while our results and other studies about visuoconstruction have demonstrated the VVT to be a valid diagnostic tool for the detection of AD (Valencia & Lehrner, 2018; Tokaj & Lehrner, 2021; Knechtl & Lehrner, 2023), the ability of the VVT 3.0 screening alone to make finer distinctions within the nonAD continuum, particularly among HC, SCD and MCI remains to be determined.
Nonetheless, the VVT 3.0 screening (delayed recall) may still be useful in the diagnosis of predementia stages of AD when used alongside other neuropsychological tests or before more expensive, complex neuropsychological testing procedures or invasive methods of neuropathological examination follow.

Longitudinal analyses

The sample size at E2 was quite small, which is why Kruskal-Wallis analyses for examining differences in delayed recall screening scores between E1 and E2 did not reveal significant results. Therefore, we concluded that there are no significant differences in scores between E1 and E2 across the diagnostic groups.
Further analysis using the Mann-Whitney U test to compare progressors and non-progressors within the diagnostic groups also failed to produce significant findings, leading us to conclude that there is no significant difference in VVT 3.0 delayed recall screening scores between progressors and non-progressors at E1 for these groups. This aligns with the research Knechtl and Lehrner conducted in 2021 for the full version of the VVT 3.0 and as in their research can be attributed to the small sample size. For future research it would be of advantage to conduct longitudinal analyses with a bigger sample size. Additionally, it is recommended to aim for repeated assessments over an extended period of time from the

initial to the final examination (Dubois et al., 2016). As Valencia and Lehrner (2018) already proposed in prior research, the time interval of 26 months in this study may be too short to adequately capture a significant number of patients transitioning from one stage to another, since it is suggested that AD evolves over many years, or possibly decades, before becoming clinically evident. Such an approach would help determine the VVT's effectiveness in tracking cognitive changes and could also offer greater insight into the progression of impairments across various cognitive domains at different stages of cognitive decline associated with AD.

Associations

In the diseased groups, we found low to moderate correlations between age and VVT 3.0 Screening scores, whereas the correlation with the delayed recall and the DI-quotient was higher than the correlation with the copy score. The negative correlations suggest that increasing age was associated with lower VVT 3.0 screening scores, which is not surprising since age is a risk factor for dementia. This is consistent with prior findings (Knechtl and Lehrner, 2023; Valencia and Lehrner, 2018; Numrich, 2017). The higher correlations observed with the delayed recall and the DI-Quotient could be attributed to the fact that the delayed recall task requires a broader range of cognitive functions, including memory and executive processes, which might decline due to AD, but also tend to decline in normal aging (Karr et al., 2018). In the healthy control group, the associations between age and the VVT 3.0 screening scores are weaker than in the diseased groups, with the strongest relationship being between age and copy scores. This implies, that the impact of aging on visuoconstructional and memory functions is less pronounced in the absence of disease-related cognitive impairment. The differences in correlation coefficients between the healthy control group and the diseased groups are relatively small for the copy scores but more noticeable for the delayed recall and DI-quotient scores. The moderate correlations in the diseased group for delayed recall and DI-quotient suggest that these cognitive domains are more affected by aging in the presence of disease than by age related decline.

All three VVT 3.0 screening scores showed moderate positive correlations with years of education, suggesting that participants with more years of education tend to perform better on the VVT 3.0 screening. This is consistent with the existing assumption that education is known to act as a protective factor (Hersi et al., 2017; Solomon et al., 2014). Given this, the inclusion of adjustments for age and education within the VVT 3.0 Screening scores, or

alternatively employing age- and education-adjusted normative values for the interpretation of results might be debated.

As anticipated, the MMSE showed strong correlations with all three VVT 3.0 screening scores, which aligns with our expectations given that the MMSE is a comprehensive cognitive assessment and includes a visuoconstructional element with its pentagon drawing task (Folstein, 1975). The stronger correlation observed with the delayed recall score and DI-quotient compared to the copy score can likely be attributed to the additional cognitive domains engaged by the delayed recall task. Therefore, it can be concluded that the VVT 3.0 screening scores are a helpful indicator of patients' overall cognitive status, which aligns with previous research on different visuoconstructive assessment tools (Cormack et al., 2004).

The correlations between the WST and the VVT 3.0 screening scores, although statistically significant, are weak to moderate, implying that cognitive intelligence has a small effect on the cognitive abilities measured by the VVT 3.0. The non-significant correlations with the BDI-II suggest that the depressive symptoms - at least as measured by the BDI-II - do not appear to have a significant impact on the cognitive domains assessed by the VVT 3.0 screening in this sample. A reason for this could be that depressive symptoms do not influence these cognitive tasks or that the effect is too small to be detected in this sample. Nonetheless, this result corresponds with prior findings (Knechtl and Lehrner, 2021; Valencia and Lehrner, 2018).

All three VVT 3.0 screening scores showed significant small to moderate associations (- .55 to .656) with almost all domains of the NTBV in the small to medium range. Overall, the correlation analysis with the NTBV indicated that the VVT 3.0 delayed recall screening task was moderately influenced by several cognitive areas, including attention, psychomotor speed, language, and executive functions. When compared to the copy task, the delayed recall demonstrates a stronger correlation with all the NTBV subtests, particularly those assessing memory. This aligns with our understanding that the delayed recall task involves episodic memory (Caruso et al., 2020). However, other than expected in the domain executive functioning, correlations between subtests and the delayed recall screening scores were not significantly higher than for the copy screening scores. For nearly all NTBV subtests, the correlation coefficient for the DI-quotient lay between the values for the delayed recall and the copy task. These findings are similar to the ones reported by Knechtl and Lehrner (2021) in their research about the full version of the VVT 3.0 delayed recall.

Correlation analyses between the DI-Quotient and the VVT 3.0 delayed recall screening score revealed a strong correlation. This is also consistent with the findings observed in our

statistical analyses, showing that the results of the DI-quotient almost always resembled the results of the delayed recall considerably stronger than the copy task. This might be due to the nature of the DI-quotient since it's calculated by dividing the delayed recall score by the copy task score, inherently linking the DI-Quotient more closely with the delayed recall performance. Since the findings for the DI-quotient typically fell between those for the copy task and the delayed recall, the quotient does not provide additional informative value beyond what is already offered by the initial screening scores of the copy task and delayed recall. This is consistent with Knechtl's (2021) finding about the DI-Quotient for the full version of the VVT 3.0. Therefore, we propose that future studies on the VVT 3.0 should eliminate the extra step of computing the DI-quotient.

The significant moderate correlation of r=.507 between the copy screening score and the delayed recall screening score suggests that, while related, the two measures reflect distinct constructs.

Drop out bias

When exploring possible drop-out bias we found no significant differences between Dropouts and follow-up participants except for the MMSE scores: Follow-up participants scored significantly higher in the MMSE at E1 than dropouts. Therefore, it seemed that the follow-up participants showed less cognitive decline as the dropouts. Further comparisons of follow-ups and dropouts regarding the three VVT 3.0 screening scores, age and education did not reach significance, presumably due to the small size of our sample. In future research, performing a drop-out analysis with a larger group and collecting data on the reasons for dropping out could prove beneficial.

6.2. Present Limitations and Suggestions on Future Research

Our study's design carries certain limitations that warrant discussion. Firstly, the severity of dementia influences our findings, particularly when diagnoses are categorized into broad categories such as nonAD or AD. Including more patients with advanced dementia would likely increase sensitivity and reduce the likelihood of false positives results. Secondly, comorbidities or conditions mimicking AD symptoms could skew test results. Participants who have AD or are in the early stages of cognitive decline leading up to AD, often have multiple health issues that make it difficult to diagnose their condition accurately (DSM-5, APA, 2013; Livingstone et al., 2020; Stephan et al., 2011). Furthermore confounders, such as apathy, which manifests as reduced motivation or initiative—a characteristic often seen in

AD —were not accounted for and could hypothetically lower scores in AD patients, resulting in an overestimation of VVT 3.0 screening scores (McKhann et al., 2011). Cognitive declines also occur in normal aging e.g. processing speed, complex attention, or fine motor skills and might also have influenced the results (Hoogendam et al., 2014). These age-related effects, whether they impact the HC or diseased group, could lead to misestimation of the study's results. The analyses in our study were based on observed raw scores and not adjusted for variables such as age or gender, which can have possible confounding effects. Future research might benefit from adjusting the VVT 3.0 Screening scores for these variables or employing age- and education-corrected normative values to interpret the results more accurately. Another limitation is sample selection bias, as our patients either sought medical advice independently or through their social networks, which potentially resulted in a sample biased towards healthier or more socially integrated individuals. Lastly the longitudinal analysis of our study was conducted with a small sample size, raising concerns about drop-out bias. Our analysis indicates that participants who only attended one examination had lower global cognitive functioning. Therefore, it's plausible that those with more severe disease progression did not return for follow-up, possibly due to poor overall health or lack of awareness of their condition's severity. These potential biases serve as a reminder to be cautious in generalizing our results too broadly. Despite these challenges, we believe our work has provided valuable insights into visuoconstructive impairments and supports the utility of the VVT screening (delayed recall) in evaluating such deficits.

6.3. Conclusion

In summary, the introduction of a delayed recall component improved the ability of the VVT 3.0. screening to distinguish between different diagnostic categories and enhance its overall diagnostic precision. Specifically, the delayed recall screening scores were able to differentiate between the diagnostic groups, except between HC - SCD and SCD - naMCI. Moreover, the delayed recall screening demonstrated high diagnostic accuracy, evidenced by an AUC of .898 and a corresponding sensitivity of 85.6% and specificity of 82% at a cut-off of 2.5 points. According to the regression model 84.7% of cases could be classified correctly as AD versus nonAD. All values are surpassing the performance of the immediate copy task, which is likely due to the delayed recall's additional reliance on memory.

While those findings are very satisfying within our study sample, they should be approached with caution when considering their applicability to other groups, due to potential biases. Consequently, further research involving diverse samples is essential to validate its

effectiveness in clinical diagnosis, disease staging, and cognitive assessment. While the VVT 3.0 screening shows good diagnostic values, it should serve as an additional test in the diagnostic process by providing information about the cognitive status and used alongside other diagnostic tools and clinical assessments to provide a comprehensive evaluation.
The VVT 3.0 screening could also prove valuable for patient monitoring and prognosis, but this requires more comprehensive longitudinal studies with larger follow-up samples and multiple follow-up points to fully explore its utility.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental Disorders (5th ed.) https://doi.org/10.1176/appi.books.9780890425596

Amodeo, S., et al. (2014). The times they are a-changin': Clock drawing and prediction of dementia. *Journal of Geriatric Psychiatry and Neurology, 28*(2), 145–155. https://doi.org/10.1177/0891988714554709

Aisen, P. S., Petersen, R. C., Donohue, M., & Weiner, M. W. (2015). Alzheimer's disease neuroimaging initiative 2 clinical core: Progress and plans. *Alzheimer's & Dementia, 11*(7), 734-739. https://doi.org/10.1016/j.jalz.2015.05.005

Albert, M. S., DeKosky, S. T., Dickson, D., et al. (2011). The diagnosis of mild cognitive impairment 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), 270-279. https://doi.org/10.1016/j.jalz.2011.03.008

Blennow, K., Hampel, H., Weiner, M., & Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology, 6*, 131–144.

Boyle, P. A., Wilson, R. S., Yu, L., et al. (2013). Much of late life cognitive decline is not due to common neurodegenerative pathologies. *Annals of Neurology, 74*, 478–489.

Bondi, M. W., Salmon, D. P., & Kaszniak, A. W. (2009). The neuropsychology of dementia. In I. Grant & K. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders* (pp. 159–198). Oxford University Press.

Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica, 82*(4), 239-259.

Breton, A., Casey, D., & Arnaoutoglou, N. A. (2019). Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *International Journal of Geriatric Psychiatry, 34*(2), 233–242. https://doi.org/10.1002/gps.5016

Buccione, I., Perri, R., Carlesimo, G. A., Fadda, L., Serra, L., Scalmana, S., & Caltagirone, C. (2007). Cognitive and behavioural predictors of progression rates in Alzheimer's disease. *European Journal of Neurology, 14*(4), 440–446.

https://doi.org/10.1111/j.1468-1331.2007.01693.x

Caruso, G., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2020). Recall and recognition in Alzheimer's disease and frontotemporal dementia. *Journal of Alzheimer's Disease, 77*, 655-666.

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. https://doi.org/10.1212/wnl.55.12.1847

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.) Lawrence Erlbaum Associates, Inc.

Cormack, F., Aarsland, D., Ballard, C., & Tovée, M. J. (2004). Pentagon drawing and neuropsychological performance in Dementia with Lewy Bodies, Alzheimer's disease, Parkinson's disease, and Parkinson's disease with dementia. International Journal of Geriatric Psychiatry, 19(4), 371–377. https://doi.org/10.1002/gps.1094

Curtis, C., Gamez, J. E., Singh, U., et al. (2015). Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. *JAMA Neurology*, 72.

Deuschl, G., Maier, W., et al. (2016). S3 Leitlinie Demenzen [S3 guidelines on dementia]. *In Leitlinien für Diagnostik und Therapie in der Neurologie*. Deutsche Gesellschaft für Neurologie.

De Bruijn, R. F., Bos, M. J., Portegies, M. L., et al. (2015). The potential for prevention of dementia across two decades: The prospective, population-based Rotterdam Study. *BMC Medicine, 13*, 132.

Dubois, B., Feldman, H. H., Jacova, C., et al. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *The Lancet Neurology, 13*(6), 614-629. https://doi.org/10.1016/s1474-4422(14)70090-0

Duits, F. H., Prins, N. D., Lemstra, A. W., et al. (2015). Diagnostic impact of CSF biomarkers for Alzheimer's disease in a tertiary memory clinic. *Alzheimer's & Dementia, 11*, 523–32.

De Jager, C. A., Hogervorst, E., Combrinck, M., & Budge, M. M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine, 33*, 1039–1050.

Ehreke, L., et al. (2010). Is the Clock Drawing Test a screening tool for the diagnosis of mild

cognitive impairment? A systematic review. *International Psychogeriatrics, 22*(1), 56–63. https://doi.org/10.1017/S1041610209990676

Ewald, B. (2006). Post hoc choice of cut points introduced bias to diagnostic research. *Journal of Clinical Epidemiology, 59*(8), 798-801.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Francis, P. T., Ramírez, M. J., & Lai, M. K. (2010). Neurochemical basis for symptomatic treatment of Alzheimer's disease. Neuropharmacology, 59, 221–229. https://doi.org/10.1016/j.neuropharm.2010.02.010

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. I., & Delis, D. C. (1994). *Clock drawing: A neuropsychological analysis*. Oxford University Press.

Gatterer, G. (2008). *Alters-Konzentrations-Test (AKT)* (2., neu normierte Aufl.). Hogrefe.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., et al. (2006). Mild cognitive impairment. *Lancet, 367*(9518), 1262–1270.

Gessner & Arndt (2019a). Likelihood Ratio, positives. In Lexikon der Medizinischen Laboratoriumsdiagnostik. Retrieved October 04, 2023, from https://link.springer.com/chapter/10.1007/978-3-662-48986-4_1874

Gessner & Arndt (2019b). Likelihood Ratio, negatives. In Lexikon der Medizinischen Laboratoriumsdiagnostik. Retrieved October 04, 2023, from https://link.springer.com/chapter/10.1007/978-3-662-48986-4_1873.

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

Hansen, N., Hirschel, S., Teegen, B., Wiltfang, J., & Malchow, B. (2022). Preserved visuoconstruction in patients with Alzheimer's pathology and anti-neural autoantibodies: A case control study. *Frontiers in Dementia, 1*, 975851.

Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science, 297*(5580), 353–356.

Hautzinger, M., Keller, F., & Kühner, C. (2009). *BDI-II. Beck-Depressions-Inventar*. Revision (2. Aufl.). Pearson Assessment.

Hoogendam, Y. Y., Hofman, A., Van der Geest, J. N., Van der Lugt, A., & Ikram, M. A. (2014). Patterns of cognitive function in aging: The Rotterdam study. European Journal of Epidemiology, 29(2), 133-140. https://doi.org/10.1007/s10654-014-9885-4

Hersi, M., Irvine, B., Gupta, P., Gomes, J., Birkett, N., & Krewski, D. (2017). Risk factors

associated with the onset and progression of Alzheimer's disease: A systematic review of the evidence. *Neurotoxicology, 61*, 143–187. https://doi.org/10.1016/j.neuro.2017.03.006

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., et al. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia, 8*(1), 1–13. https://doi.org/10.1016/j.jalz.2011.10.007

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Brill-Edwards, P., Browman, G., Cook, D., Farkouh, M., Gerstein, H., Haynes, B., Hayward, R., Holbrook, A., Juniper, E., Lee, H., Levine, M., Moyer, V., Nishikawa, J., Oxman, A., ... Wilson, M. (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. https://doi.org/10.1001/jama.1994.03510330081039

Jessen, F., Wiese, B., Bachmann, C., Eifflaender-Gorfer, S., Haller, F., Kölsch, H., et al. (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., Spottke, A., Boecker, H., Brosseron, F., Buerger, K., Catak, C., Fliessbach, K., Franke, C., Fuentes, M., Heneka, M. T., Janowitz, D., Kilimann, I., Laske, C., Menne, F., Nestor, P., Peters, O., Priller, J., Pross, V., Ramirez, A., ... Duzel, E. (2018). Design and first baseline data of the DZNE multicenter observational study on p redementia Alzheimer's disease (DELCODE). *Alzheimer´s Research and Therapy, 10*(1). https://doi.org/10.1186/s13195-017-0314-2 10.1186/s13195-017-0314-2

Jessen, F., Amariglio, R. E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., Van Der Flier, W. M., Glodzik, L., van Harten, A. C., de Leon, M. J., McHugh, P., Mielke, M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., ... Wagner, M. (2014). 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., Amariglio, R. E., Buckley, R. F., et al. (2020). The characterisation of subjective cognitive decline. *Lancet Neurology, 19*(3), 271-278. https://doi.org/10.1016/s14744422(19)30368-0

Johnson, D. K., et al. (2009). Longitudinal study of the transition from healthy aging to Alzheimer disease. Archives of Neurology, 66(10), 1254–1259.

https://doi.org/10.1001/archneurol.2009.158

Karr, J. E., et al. (2018). When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. Psychology and Aging, 33(2), 195–218. https://doi.org/10.1037/pag0000236

Karran, E., Mercken, M., & De Strooper, B. (2011). The amyloid cascade hypothesis for Alzheimer's disease: An appraisal for the development of therapeutics. Nature Reviews Drug Discovery, 10, 698–712.

Kovacs, G. G., Milenkovic, I., Wôhrer, A., et al. (2013). Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: A community-based autopsy series. *Acta Neuropathologica*, *126*, 365–384.

Knechtl, P., & Lehrner, J. (2023). Visuoconstructional abilities of patients with subjective cognitive decline, mild cognitive impairment and Alzheimer's disease. Journal of Geriatric Psychiatry and Neurology. https://doi.org/10.1177/08919887221135549

Kumar, A., Singh, A., & Ekavali. (2015). A review on Alzheimer's disease pathophysiology and its management: An update. Pharmacological Reports, 67, 195–203. https://doi.org/10.1016/j.pharep.2014.09.004

Landis, J. R., & Koch, G. G. (1997). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*, *396*(10248), 413-446.

Leeflang, M. M., Moons, K. G., Reitsma, J. B., & Zwinderman, A. H. (2008). Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. *Clinical Chemistry*, *54*(4), 729-737.

Lehrl, S., & Fischer, B. (1997). Kurztest für cerebrale Insuffizienz (c.I.-Test) (5. überarb. Aufl.). Vless.

Lehrner, J. (2015). Vienna Visuo-konstruktive Test (VVT 3.0) – ein Verfahren zur Bestimmung der Visuokontruktion. Medical University of Vienna, Unpublished Manual.

Lehrner, J., Kogler, S., Lamm, C., Moser D., Klug, S., Pusswald, G., Dal-Bianco, P., Pirker, W., & Auff, E. (2015). Awareness of memory deficits in subjective cognitive decline, mild cognitive impairment, Alzheimer´s disease and Parkinson´s disease. International

Psychogeriatrics, 27(3), 357-366. https://doi.org/10.1017/S1041610214002245

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. https://doi.org/10.1024/1016-264X/a000175

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., Gufler, R., Guttmann, G., Maly, J., Gleiß, A., Auff, E., et al. (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

Lemeshow, S., Sturdivant, R. X., & Hosmer, D. W. (2013). *Applied Logistic Regression.* Wiley.

Linden, A. (2006). Measuring diagnostic and predictive accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis. *Journal of Evaluation in Clinical Practice, 12*(2), 132-139. https://doi.org/10.1111/j.1365-2753.2005.00598.x

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test *Assessment. Journal of Thoracic Oncology*, *5*(9), 1315-1316.

Mendez, M. F., Ala, T., & Underwood, K. L. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *Journal of the American Geriatrics Society, 40*(11), 1095-1099.

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. https://doi.org/10.1371/journal.pone.0038268

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly J. J., Mayeux, R., Mohs, J. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (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: the journal of*

the Alzheimer's Association, *7*(3), 263-269. https://doi.org/10.1016/j.jalz.2011.03.005

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

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699.

Numrich, J. (2017). *Visuo-constructional functions in patients with subjective cognitive decline (SCD), mild cognitive impairment (MCI), Alzheimer's disease (AD) and Parkinson's disease (PD)* [Unpublished master's thesis]. Medical University of Vienna.

Nestor PJ, Scheltens P, Hodges JR. Advances in the early detection of Alzheimer's disease. 2004.

Norton, S., Matthews, F. E., Barnes, D. E., Yaffe, K., & Brayne, C. (2014). Potential for primary prevention of Alzheimer's disease: An analysis of population-based data. *Lancet Neurology, 13*, 788–94.

Oswald, W. D., & Fleischmann, U. M. (1997). *Das NürnbergerAlters-Inventar*. Hogrefe.

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., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., Boeve, B. F., Tangalos, E. G., Ivnik, R. J., & Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men: The Mayo Clinic Study of Aging. *Neurology*, *75*(10), 889-897. https://doi.org/10.1212/WNL.0b013e3181f11d85

Reitan, R. M. (1979). *Trail Making Test: Manual for Administration and Scoring*. Reitan Neuropsychology Laboratory.

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

Segal, D. L., Coolidge, F. L., Cahill, B. S., & O´Riley, A. A. (2008). Psychometric properties of the Beck Depression Inventory-II (BDI-II) among community-dwelling older adults. *Behavior Modification*, *32*(1), 3-20. https://doi.org/10.1177/0145445507303833

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., et al. (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals*

of Neurology, 65(4), 403–413.

Shon, J. M., Lee, D. Y., Seo, E. H., Sohn, B. K., Kim, J. W., Park, S. Y., et al. (2013). Functional neuroanatomical correlates of the executive clock drawing task (CLOX) performance in Alzheimer's disease: A FDG-PET study. *Neuroscience, 246*, 271-280.

Stephan, B. C., Brayne, C., Savva, G. M., & Matthews, F. E. (2011). Occurrence of medical co-morbidity in mild cognitive impairment: Implications for generalisation of MCI research. Age and Ageing, 40(4), 501-507. https://doi.org/10.1093/ageing/afr057

Small, S. A., & Duff, K. (2008). Linking Aβ and tau in late-onset Alzheimer's disease: A dual pathway hypothesis. *Neuron, 60*(6), 534–542.

Solomon, A., Mangialasche, F., Richard, E., Andrieu, S., Bennett, D. A., Breteler, M., Fratiglioni, L., Hooshmand, B., Khachaturian, A. S., Schneider, L. S., Skoog, I., & Kivipelto, M. (2014). Advances in the prevention of Alzheimer's disease and dementia. Journal of Internal Medicine, 275, 229–250. https://doi.org/10.1111/joim.12178

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

Tsoi, K. K. F., et al. (2015). Cognitive tests to detect dementia: A systematic review and meta-analysis. *JAMA Internal Medicine, 175*(9), 1450–1458. https://doi.org/10.1001/jamainternmed.2015.2152

Tokaj, A., & Lehrner, J. (2023). Drawing a line? Visuo-constructive function as discriminator between healthy individuals, subjective cognitive decline, mild cognitive impairment and Alzheimer's disease and predictor of disease progress compared to a multimodal approach. *neuropsychiatrie*, 1-11.

Thomann, P. A., et al. (2008). Clock drawing performance and brain morphology in mild cognitive impairment and Alzheimer's disease. *Brain and Cognition, 67*(1), 88–93. https://doi.org/10.1007/978-1-4939-7647-8_14

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. https://doi.org/10.1007/s40211-018-0279-9.

Van Harten, A. C., Mielke, M. M., Swenson-Dravis, D. M., et al. (2018). Subjective cognitive decline and risk of MCI. *Neurology, 91*(4), e300-e312. https://doi.org/10.1212/wnl.0000000000005863

Visser, P. J., Verhey, F., Knol, D. L., et al. (2009). Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive

impairment or mild cognitive impairment in the DESCRIPA study: A prospective cohort study. *Lancet Neurology, 8*, 619–627.

Wang, D. S., Dickson, D. W., & Malter, J. S. (2006). β-Amyloid degradation and Alzheimer's disease. *Journal of Biomedical Biotechnology, 2006*, Article 58406. https://doi.org/10.1155/JBB/2006/58406

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine, 2*(4), a006171. https://doi.org/10.1101/cshperspect.a006171

Wirtz, M. A. (2019a). Negative Predictive Value. In Dorsch Lexikon der Psychologie. Retrieved October 04, 2023, from https://dorsch.hogrefe.com/stichwort/negativ-praediktiver-wert

Wirtz, M. A. (2019b). Positive Predictive Value. In Dorsch Lexikon der Psychologie. Retrieved October 04, 2023, from https://dorsch.hogrefe.com/stichwort/positiv-praediktiver- wert-ppv

Wirtz, M. A. (2019c). Sensitivität. In Dorsch Lexikon der Psychologie. Retrieved October 05, 2023, from https://dorsch.hogrefe.com/stichwort/sensitivitaet

Wirtz, M. A. (2019c). Specificity. In Dorsch Lexikon der Psychologie. Retrieved October 05, 2023, from https://dorsch.hogrefe.com/stichwort/spezifitaet

Wolfsgruber, S., Polcher, A., Koppara, A., Kleineidam, L., Frölich, L., Peters, O., Hüll, M., Rüther, E., Wiltfang, J., Maier, W., Kornhuber, J., Lewczuk, P., Jessen, F. & Wagner, M. (2017). Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment. *Journal of Alzheimers Disease, 58*(3), 939–50. https://doi.org/10.3233/JAD-161252

Youden, W. J. (1950). Index for rating diagnostic rests. *ACS Journals, 3*(1), 32-53. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zec, R. F., Landreth, E. S., Vicari, S. K., Belman, J., Feldman, E., Andrise, A., et al. (1992). Alzheimer Disease Assessment Scale: A subtest analysis. *Alzheimer Disease and Associated Disorders, 6*(3), 164-181.

List of Abbreviations

AD – Alzheimer's disease

AMCI – amnestic mild cognitive impairment

APA – American psychiatric association

AUC – Area under the curve

BDI-II - Beck Depression Inventory-II

CI – confidence interval

E1/2 – examination 1/2

HC – healthy control

IQ – intelligence quotient

IQR – interquartile range

DI-quotient – delayed recall/immediate copy-quotient

MCI – mild cognitive impairment

MMSE - Mini-Mental State Examination

MOCA - Montreal Cognitive Assessment

N/n – absolute number

NaMCI – non amnestic mild cognitive impairment

NCD – neurocognitive disorder

NTBV – Neuropsychological Test Battery Vienna

ROC - receiver operating characteristic

SCD – subjective cognitive decline

SD – standard deviation

VCD – Vienna conversion to dementia

VVT - Vienna Visuoconstructional Test

List of Tables

List of Figures

Appendix

Abstract

Background: For distinguishing between normal cognition and different stages of cognitive impairment caused by Alzheimer's disease (AD), the growing consensus suggests the assessment of various cognitive domains that are impaired during the disease. Visuoconstructional functions, executive functions and memory are some of the various cognitive abilities that are impaired within the course of AD. We assessed those cognitive abilities with the Screening version of the Vienna Visuoconstructional Test 3.0 (VVT 3.0) Delayed recall and investigated the VVT's potential as a diagnostic screening tool.

Methods: We retrospectively analyzed the data of 640 patients, that were classified into the following diagnostic groups: 150 healthy controls (HC), 49 with subjective cognitive decline (SCD), 114 with amnestic mild cognitive impairment (aMCI), 166 with non amnestic mild cognitive impairment (naMCI) and 161 with AD. Using Kruskal-Wallis and Dunn's post hoc comparison analyses we investigated the differences in the VVT 3.0 screening scores. Receiver operating characteristic (ROC) curves were conducted, to evaluate the ability of the VVT 3.0 screening to differentiate between AD and nonAD, by defining ideal cut-offs. In addition, a binary logistic regression model was computed. Furthermore, we conducted a follow-up with 40 of our initial 640 participants 12-48 months later to capture alterations in performance over time. At last, we studied correlations between various clinical variables, the VVT 3.0 screening scores and longitudinal performance changes.

Results: Results stated that the VVT 3.0 delayed recall screening scores were able to differentiate between all diagnostic groups, respectively, except HC – SCD and SCD - aMCI. The ROC analyses determined an AUC of 0.898, 95% CI [0.873;0.923], $p<0.001$. The ideal cut-off using the Youden Index was determined at 2.5 points, with a corresponding sensitivity at 0.856 and specificity at 0.820. The logistic regression model classified 84.5% of AD patients correctly and delivered a significant Cohen's Kappa of 0.621 ($p<0.001$). Longitudinal analyses illustrated improving, consistent as well as deteriorating courses of disease. The remaining longitudinal analyses, however, did not reach significance.

Conclusion: As the screening version of the VVT 3.0 demonstrated good diagnostic accuracy in our sample, it could support clinical diagnosing of the preclinical and prodromal stages due to AD and help in specifying the cognitive status. Nevertheless, to better

understand its utility across different populations and over extended periods, there is a necessity for further research involving diverse samples and with more prolonged follow-up periods.
Keywords: *visuoconstruction, memory, executive functions, delayed recall, screening, subjective cognitive decline, mild cognitive impairment, alzheimer's disease*

Abstract in German / Abstract in Deutsch

Hintergrund: Um zwischen normaler Kognition und verschiedenen Stadien der kognitiven Beeinträchtigung durch die Alzheimer-Krankheit (AD) zu unterscheiden, schlägt der wachsende Konsens die Bewertung verschiedener kognitiver Bereiche vor, die während der Krankheit beeinträchtigt werden. Zu diesen gehören visuell-konstruktive Funktionen, exekutive Funktionen und das Gedächtnis. Wir untersuchten diese kognitiven Fähigkeiten mit der Screening-Version des Wiener Visuokonstruktionstests 3.0 (VVT 3.0; Verzögerter Abruf) und untersuchten das Potenzial des VVT als diagnostisches Screening-Instrument.
Methoden: Wir analysierten retrospektiv die Daten von 640 Patienten, die in die folgenden diagnostischen Gruppen eingeteilt wurden: 150 gesunde Kontrollen (HC), 49 mit subjektivem kognitivem Verfall (SCD), 114 mit amnestischer leichter kognitiver Beeinträchtigung (aMCI), 166 mit nicht amnestischer leichter kognitiver Beeinträchtigung (naMCI) und 161 mit Alzheimer. Anhand von Kruskal-Wallis- und Dunn's Post-Hoc-Vergleichsanalysen untersuchten wir die Unterschiede in den VVT 3.0-Screening-Scores. Es wurden Receiver-Operating-Characteristic (ROC)-Kurven erstellt, um die Fähigkeit des VVT 3.0-Screenings zur Unterscheidung zwischen Alzheimer und Nicht-Alzheimer zu bewerten, indem ideale Cut-offs definiert wurden. Darüber hinaus wurde ein binäres logistisches Regressionsmodell berechnet. Zuletzt führten wir mit 40 unserer 640 Teilnehmer eine Nachuntersuchung 12-48 Monate später durch, um Veränderungen der Leistung im Laufe der Zeit zu erfassen und schließlich untersuchten wir Korrelationen zwischen verschiedenen klinischen Variablen, den VVT 3.0-Screening-Scores und longitudinalen Leistungsänderungen.
Ergebnisse: Die Ergebnisse zeigten, dass die VVT 3.0-Screening-Scores (Verzögerter Abruf) in der Lage waren, zwischen allen diagnostischen Gruppen zu unterscheiden, mit Ausnahme von HC - SCD und SCD - aMCI. Die ROC-Analysen ergaben eine AUC von 0,898, 95% CI [0,873;0,923], $p<0,001$. Der ideale Cut-off-Wert unter Verwendung des Youden-Index lag bei 2,5 Punkten, mit einer entsprechenden Sensitivität von .856 und einer Spezifität von .820. Das logistische Regressionsmodell klassifizierte 84,5 % der AD-Patienten korrekt und lieferte ein signifikantes Cohen's Kappa von .621 ($p<.001$). Die

Längsschnittanalysen zeigten sowohl einen sich verbessernden, gleichbleibenden als auch einen sich verschlechternden Verlauf der Krankheit. Die übrigen Längsschnittanalysen erreichten jedoch keine Signifikanz.

Schlussfolgerung: Da die Screening-Version des VVT 3.0 in unserer Stichprobe eine gute diagnostische Genauigkeit aufwies, könnte sie die klinische Diagnose der präklinischen und prodromalen Stadien der Alzheimer-Krankheit unterstützen und bei der Spezifizierung des kognitiven Status helfen. Um die Nützlichkeit des Tests in verschiedenen Populationen und über längere Zeiträume hinweg besser zu verstehen, sind jedoch weitere Untersuchungen mit unterschiedlichen Stichproben und mit längeren Nachbeobachtungszeiträumen erforderlich.

Schlüsselwörter: *Visuokonstruktion, Gedächtnis, exekutive Funktionen, verzögerter Abruf, Screening, subjektiver kognitiver Verfall, leichte kognitive Beeinträchtigung, Alzheimer-Krankheit*