

Visual associative learning in Alzheimer's Disease and performance validity

Citation for published version (APA):

Meyer, S. (2020). Visual associative learning in Alzheimer's Disease and performance validity: New applications of the Visual Association Test. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20200911sm>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20200911sm](https://doi.org/10.26481/dis.20200911sm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**VISUAL ASSOCIATIVE LEARNING IN
ALZHEIMER'S DISEASE AND
PERFORMANCE VALIDITY
NEW APPLICATIONS OF THE VISUAL
ASSOCIATION TEST**

SASCHA R.A. MEYER

**VISUAL ASSOCIATIVE LEARNING IN ALZHEIMER'S DISEASE AND
PERFORMANCE VALIDITY**

New applications of the Visual Association Test

Sascha R.A. Meyer

Copyright © S.R.A. Meyer, Maastricht, 2020

Cover design: Sascha R.A. Meyer

Lay out: ProefschriftMaken || www.proefschriftmaken.nl

Printed by: ProefschriftMaken || www.proefschriftmaken.nl

ISBN: 978 94 6380 879 8

**VISUAL ASSOCIATIVE LEARNING IN ALZHEIMER'S DISEASE AND
PERFORMANCE VALIDITY**

New applications of the Visual Association Test

PROEFSCHRIFT

Ter verkrijging van de graad van doctor

aan de Universiteit Maastricht

op gezag van Rector Magnificus, Prof. dr. Rianne M. Letschert,

volgens het besluit van het College van Decanen

in het openbaar te verdedigen op

vrijdag 11 september 2020 om 12:00 uur

door

Sascha Rainer Albert Meyer

Geboren op 21 mei 1970 te Hamburg (Duitsland)

Promotor:

Prof. dr. R.W.H.M. Ponds

Co-promotor:

Dr. J.F.M. de Jonghe (Northwest Medical Center Alkmaar)

Beoordelingscommissie:

Prof. dr. F.R.J. Verhey (voorzitter)

Prof. dr. M. Jelicic

Prof. dr. R.P.C. Kessels (Radboud UMC Nijmegen)

Prof. dr. E.J.A. Scherder (VU Amsterdam)

The research described in this thesis was performed at the department of geriatric medicine of the Northwest Medical Center Alkmaar, the Netherlands, and the neuropsychological practice of clinical neuropsychologist Dr. J.F.M. de Jonghe.

The research presented in this thesis was supported by Foreest Northwest Medical School, Alkmaar, the Netherlands [grant number FIO1308].

Voor Annemieke, Sven en Wessel

CONTENTS

| | | |
|-----------|--|-----|
| CHAPTER 1 | General introduction | 9 |
| PART I | PRACTICAL ISSUES IN MEASURING EPISODIC MEMORY | |
| CHAPTER 2 | Visual associations cued recall: A paradigm for measuring episodic memory decline in Alzheimer's disease <i>Aging, Neuropsychology, and Cognition, 2016</i> | 23 |
| PART II | EPISODIC MEMORY IN ALZHEIMER'S DISEASE | |
| CHAPTER 3 | Episodic recognition memory based on incidental learning of visual associations is largely preserved compared to recall in amnesic mild cognitive impairment and mild Alzheimer's disease <i>Applied Neuropsychology: Adult, 2019</i> | 47 |
| CHAPTER 4 | Visual associations to retrieve episodic memory across healthy elderly, mild cognitive impairment, and patients with Alzheimer's disease <i>Aging, Neuropsychology, and Cognition, 2019</i> | 75 |
| CHAPTER 5 | Testing episodic memory in elderly subjects: Not as simple as it looks <i>Dementia and Geriatric Cognitive Disorders, 2019</i> | 107 |
| PART III | PERFORMANCE VALIDITY OF EPISODIC MEMORY | |
| CHAPTER 6 | The Visual Association Test-Extended: A cross-sectional study of the performance validity measures <i>The Clinical Neuropsychologist, 2017</i> | 131 |
| CHAPTER 7 | General discussion | 165 |
| ADDENDUM | Summary | 185 |
| | Samenvatting (Dutch summary) | 189 |
| | Knowledge valorization | 193 |
| | Author affiliations | 205 |
| | Publications | 207 |
| | Dankwoord (Acknowledgements) | 209 |
| | Curriculum Vitae | 213 |

1

Chapter 1

GENERAL INTRODUCTION



GENERAL INTRODUCTION

In this thesis we investigated visual associative learning in Alzheimer's disease (AD) and in patients who show low performance validity. For this purpose, we developed new applications of the Visual Association Test (VAT; Lindeboom & Schmand, 2003), i.e. parallel versions C and D of the VAT (Lindeboom, Schmand, Meyer, & de Jonghe, 2014), the measures of memory of the Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017; Meyer & de Jonghe, 2019), and the performance validity measures of the VAT-E. In this chapter we discuss challenges in testing episodic memory in AD, and how the VAT and the VAT-E may be used to test episodic memory from normal aging to Alzheimer's dementia and how the VAT-E may be used to test performance validity. Also, this chapter provides the aim and outline of this thesis.

1. Challenges in testing episodic memory in Alzheimer's dementia and mild memory impairment

Episodic memory deficits are the hallmark of Alzheimer's dementia. However, testing such deficits comes with several challenges, like floor effects when the disease progresses and different effects on memory performance as a function of type of retrieval support.

Floor effects

Episodic memory deficits are an early sign of Alzheimer's dementia and are best measured with neuropsychological tests that capture different memory processes. Free recall tests, such as the well-known Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964; Schmidt, 1996), adequately differentiate healthy controls from AD patients (Eckström et al., 2013; Giffort et al., 2015), but these tests pose a challenge to patients once their episodic memory impairment is moderately severe which results in floor effects (Bigler, Rosa, Schultz, Hall & Harris, 1989;

Gifford et al., 2015; Kessels, Rijken, Joosten-Weyn Banningh, Schuylenborgh-van Es, van & Olde Rikkert, 2009; Petersen et al., 1999; Woodward, Dunlosky & Salthouse, 1999). Floor effects on tests implies that measures can be clinically useful in the early stages of the disease, but soon after lose the ability to describe the remaining memory capacities.

Retrieval support

Recent findings indicate that there is more than just impaired consolidation in AD. Transgenic mouse models of early AD suggest that much of the newly learned information might actually be stored in long term memory (Roy et al., 2016), but that the way the information is retrieved is essential to whether it is remembered or not (Roy et al., 2016). If clinicians better knew what memory capacities remain and, more specifically, how they function, health care could be aligned more effectively with the needs of patients.

One of the major challenges is to investigate to what extent amnesic mild cognitive impairment (a-MCI) patients and AD patients can remember newly learned information under different types of retrieval support. Retrieval support refers to ways that aid patients, during testing, to better remember learned information. This may be done by means of presenting information related to it that was encoded along with the target stimulus, i.e. a cue, by means of presenting the learned information itself, i.e. recognition (Tulving & Madigan, 1970), or by means of intensified test guidance, i.e. raising the awareness of the test administrator regarding the importance of full completion of a test and the patient being given continuous support and encouragement during testing.

2. Testing episodic memory from normal aging to Alzheimer's dementia

Few test are available that can investigate episodic memory from normal aging to severe Alzheimer's dementia, as either tests are too difficult for patients resulting in floor effects

(Bigler et al., 1989; Gifford et al., 2015; Kessels et al., 2009; Petersen et al., 1999; Woodward et al., 1999) or tests are too easy for normal subjects resulting in ceiling effects (Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002). To assess the level of intact memory performance across the episodic memory continuum from normal aging to AD dementia, we developed new applications of the VAT. The VAT is based on incidental learning of visual associations which are tested by means of cued recall, i.e. paired associate recall. First, to monitor disease progression, we developed two extra parallel versions of the VAT, i.e. versions C and D. Second, to cover a broad range of the episodic memory continuum, we developed the VAT-E by joining the items of all four VAT versions together, i.e. versions A, B, C and D, and adding free recall. Only the parallel versions C and D have a recognition trial, which we used for the multiple-choice cued recognition trial of the VAT-E. By doing so, we constructed three measures of memory of the VAT-E, i.e. paired associate recall, free recall and multiple-choice cued recognition, each of which provides a different level of retrieval support. Also, we integrated the patient being given continuous support and encouragement during testing into the VAT-E test instruction.

3. Visual Association Test-Extended and performance validity

A potential threat to memory assessment is that test results may be distorted by patients who show low performance validity. An estimated 6.5% of patients overall and 13% of patients younger than 65 years of age referred to memory clinics show such low performance validity (Rienstra et al., 2013). In addition, an estimated 22 to 40% of patients involved in litigation (Larrabee, 2007; Mittenberg, Patton, Canyock, & Condit, 2002) also show low performance validity. It is essential to detect these patients to prevent incorrect diagnoses (Bush et al., 2005; Guilmette, 2013; Heilbronner, et al., 2009) or incorrect allocations of scarce health care resources. Low performance validity refers to memory performance that is discrepant with

known patterns of memory functioning (Green, 2003; Rogers, 2008; Slick et al., 1999). Such discrepancies can be detected by means of perform validity tests, which compare the memory performance of patients who take these tests with the memory performance of patients who have a genuine memory impairment (Green, 2003; Rogers, 2008; Slick et al., 1999). Two methods are frequently used. First, multiple-choice trials of recognition show ceiling effects for patients with a genuine memory impairment, which patients who show low performance validity do not show (Green, Iverson, & Allen, 1999; Tombaugh, 1996). Second, patients who have a genuine memory impairment show improvement on multiple-choice trials of recognition compared to free recall trials, which patients who show low performance validity show to a much lesser degree (Green, 2003; Haines & Norris, 1995; Rogers, 2008; Slick et al., 1999). In line with these two methods, we developed the VAT-E performance validity measures, i.e. immediate recognition, delayed recognition, consistency between immediate and delayed recognition, and the embedded performance validity measure free recall compared to multiple-choice cued recognition.

4. Aim of the thesis

The first aim of this thesis was to investigate episodic memory functioning from normal aging to AD dementia using the VAT parallel versions C and D and the measures of memory of the VAT-E. A second aim was to examine whether the VAT-E performance validity measures could be used as a new performance validity test. More specific, we addressed the following research questions:

1. Does the VAT show a smaller floor effect in a group consisting of healthy elderly controls, a-MCI patients and mild AD patients than the RAVLT?
2. Is episodic recognition memory preserved compared to recall in a-MCI patients and mild AD patients?

3. To what extent can a-MCI patients and mild AD patients still benefit from retrieval support such as cueing or recognition when incidentally learning visual associations?
4. Does intensified test guidance on the VAT or the RAVLT reduce the number of missing values in groups of healthy elderly controls, a-MCI patients and mild AD patients and does it improve their memory performance?
5. Do the performance validity measures of the VAT-E differentiate a-MCI patients and mild AD patients from persons who were instructed to feign memory problems?
6. Do the performance validity measures of the VAT-E differentiate litigating patients who were classified as Malingering of Neurocognitive Dysfunction from litigating patients who were classified as non-Malingering of Neurocognitive Dysfunction?

5. Outline of the thesis

This thesis consists of three parts. The first part describes practical issues in measuring episodic memory in AD dementia. **Chapter 2** describes whether the RAVLT or the VAT can be used to monitor the progression of the episodic memory impairment in a- MCI and AD. In a group consisting of healthy elderly controls, a-MCI patients, and AD patients, floor effects of RAVLT immediate free recall and VAT paired associate recall were compared in relation to global cognitive impairment.

The second part describes the episodic memory functioning across the range normal aging to mild AD dementia. **Chapter 3** describes whether the episodic memory impairment in AD can be characterized as a consolidation or a retrieval impairment. The VAT-E measures of memory paired associate recall and free recall were compared with multiple-choice cued recognition within groups of a-MCI patients and mild AD patients. **Chapter 4** describes to what extent patients with moderate to severe memory deficits can still benefit from retrieval support such as cueing or recognition when incidentally learning visual associations. In a

group consisting of healthy elderly controls, a-MCI patients, and mild AD patients, score distributions of the VAT-E measures of memory paired associate recall, free recall and multiple-choice cued recognition were compared in relation to the global cognitive impairment. **Chapter 5** describes whether intensified test guidance on the VAT or the RAVLT can be used to reduce the number of missing values, and can be used to improve the memory performance of patients. The number of missing values and mean test results on the VAT and the RAVLT were compared between intensified test guidance and routine test administration in groups of healthy elderly controls, a-MCI patients, and mild AD patients.

The third part describes whether knowledge of the memory functioning across the range normal aging to mild AD dementia can be used to detect patients who show low performance validity. In **chapter 6** two studies are presented. The first study describes how the cut-off scores for VAT-E performance validity measures were established. In an experimental design, persons who were instructed to feign memory deficit were compared with healthy controls, a-MCI patients, and mild AD patients. The second study describes the usefulness of these cut-off scores to differentiate litigating patients who show low performance validity from those who do not. In a known-groups comparison, litigating patients classified as Malingering of Neurocognitive Dysfunction were compared with litigating patients classified as non-Malingering of Neurocognitive Dysfunction.

References

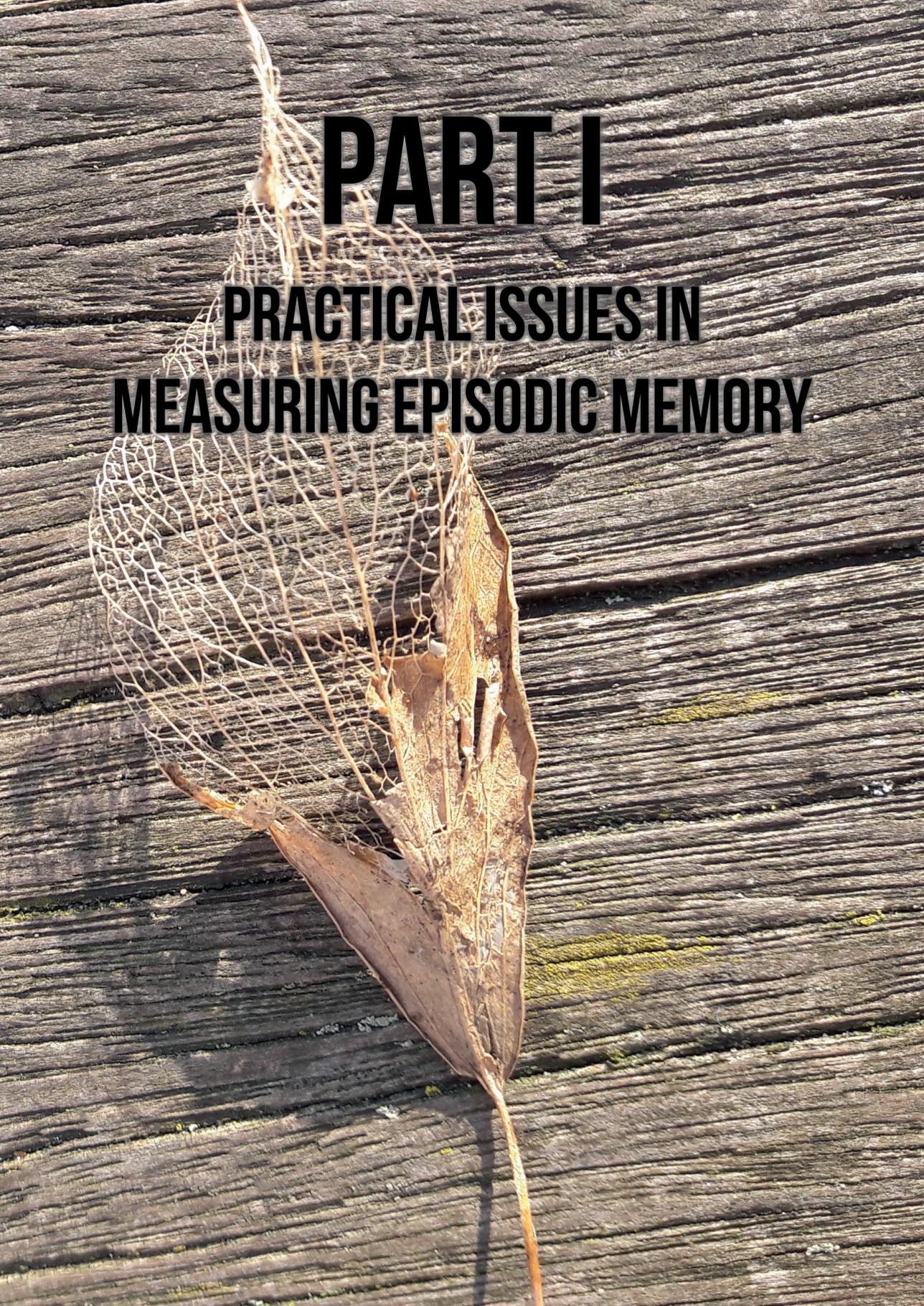
- Bigler, E. D., Rosa, L., Schultz, R., Hall, S., & Harris, J. (1989). Rey-auditory verbal learning and rey-osterrieth complex figure design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology, 45*, 277-280.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ...Silver, C.

- H. (2005). Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, 20, 419-426.
- Eckerström, C., Olsson, E., Bjerke, M., Malmgren, H., Edman, A., Wallin, A., et al. (2013). A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts conversion from mild cognitive impairment to dementia. *Journal of Alzheimer's Disease*, 36, 421-431.
- Gifford, K. A., Phillips, J. S., Samuels, L. R., Lane, E. M., Bell, S. P., Liu, D., ... for the Alzheimer's Disease Neuroimaging Initiative (2015). Associations between verbal learning slope and neuroimaging markers across the cognitive aging spectrum. *Journal of the International Neuropsychological Society*, 21, 455-467.
- Green, P. (2003). *Green's Word Memory Test for Microsoft Windows: User's manual*. Edmonton, Canada: Green's Publishing Inc.
- Green, P., Iverson, G. L., & Allen, L. (1999) Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, 13, 813-819.
- Guilmette, T. J. (2013). The role of clinical judgement in symptom validity assessment. In Carone, D. A. & Busch S. S. (Eds.) (2013), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p.p. 31-43). New York: Springer Publishing Company, LLC.
- Haines, M. E. & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology Review*, 5, 125-148.
- Heilbronner, R. L., Sweet, J. L., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093-1129.
- Kessels, R. P. C., Rijken, S., Joosten-Weyn Banningh, L. W. A., Schuylenborgh-van Es, N.

- van & Olde Rikkert, M. G. M. (2009). Categorical spatial memory in patients with mild cognitive impairment and alzheimer dementia: Positional versus object-location recall. *Journal of the International Neuropsychological Society*, *16*, 200-204.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). In Larrabee, G. J. (Ed.) (2007). *Assessment of malingered neuropsychological deficits* (p.p. 34-37). New York: Oxford University Press Inc.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test. Manual*. Leiden: PITS bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, *73*, 126-133.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual]. Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2019). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual] (2nd ed.). Amsterdam: Hogrefe Publishing bv.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094-1102.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment. *Archives of Neurology*, *56*, 303-308.
- Rey, A. (1964). *L'examen Clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Rienstra, A., Groot, P. F. C., Spaan, P. E. J., Majoie, C. B. L. M., Nederveen, A. J., Walstra,

- G. J. M., ...Schmand, B. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, 35, 59-70.
- Rogers, R. (Ed.) (2008), *Clinical assessment of malingering and deception* (3rd ed., p.p. 3-13, 14-35, 411-434). New York: The Guilford Press.
- Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature*, 531, 508-512.
- Schmidt, M. (1996). *Rey Auditory Verbal Learning Test. A handbook*. Los Angeles: Western Psychological Services.
- Slick, J. D., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standard for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545-561.
- Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, New York: Multi-Health Systems Inc.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21, 437-484.
- Woodward, J. L., Dunlosky, J., & Salthouse, T. A. (1999). Task decomposition analysis of intertrial free recall performance on the rey auditory verbal learning test in normal aging and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 21, 666-676.





PART I
PRACTICAL ISSUES IN
MEASURING EPISODIC MEMORY

2

Chapter 2

VISUAL ASSOCIATIONS CUED RECALL: A PARADIGM FOR MEASURING EPISODIC MEMORY DECLINE IN ALZHEIMER'S DISEASE

Sascha R.A. Meyer, Pauline E.J. Spaan, Leo Boelaarts, Rudolf W.H.M. Ponds,
Ben Schmand, and Jos F.M. de Jonghe

Aging, Neuropsychology, and Cognition, 2016; 23: 566-577

Abstract

Repeated measurements of episodic memory are needed for monitoring amnesic mild cognitive impairment (aMCI) and mild Alzheimer's disease (AD). Most episodic memory tests may pose a challenge to patients, even when they are in the milder stages of the disease. This cross-sectional study compared floor effects of the Visual Association Test (VAT) and the Rey Auditory Verbal Learning Test (RAVLT) in healthy elderly controls and in patients with aMCI or AD ($N = 125$). A hierarchical multiple regression analysis was used to examine whether linear or quadratic trends best fitted the data of cognitive test performance across global cognitive impairment. Results showed that VAT total scores decreased linearly across the range of global cognitive impairment, whereas RAVLT total scores showed a quadratic trend, with total scores levelling off for 90% of aMCI patients and 94% of AD patients. We conclude that the VAT shows few if any floor effects in patients with aMCI and mild AD and is therefore a potentially promising cognitive test for monitoring episodic memory impairment.

Introduction

Episodic memory impairment is the hallmark of Alzheimer's disease (AD) (Dubois et al., 2007). To closely align health care with the needs of AD patients, it is essential to monitor the disease by repeated assessment of the episodic memory impairment. For this purpose, rote repetition and word list free recall tests are available, such as the well-known Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964; Schmidt, 1996). Although these tests are highly sensitive measures for differentiating patients from healthy elderly controls (HCs), they may pose a challenge to AD patients, resulting in a diminished ability to differentiate between patients at the lower end of the measurement range (Bigler, Rosa, Schultz, Hall & Harris, 1989; Kessels, Rijken, Joosten-Weyn Banningh, Schuylenborgh-van Es, van & Olde Rikkert, 2009; Petersen et al., 1999; Woodward, Dunlosky & Salthouse, 1999). Such floor effects seriously affect the power of a test to determine the progression of the episodic memory impairment in AD (Cohen, Swerdluk & Sturman, 2013).

Determining the degree of the episodic memory impairment in AD can be improved by the use of test paradigms that provide encoding specificity, where the test material is encoded along with specific cues and with cued recall these cues are presented to maximise retrieval. When encoding specificity is provided by category cues, cued recall improves the retrieval of test material in AD compared to free recall (Grober & Buschke, 1987; Grober, Buschke, Crystal, Bang & Dresner, 1988; Pasquier, Grymonprez, Lebert & Linden, van der, 2001; Pillon, Deweer, Michon, Malapani, Agid & Dubios, 1994; Saka, Mihci, Akif Topcuoglu & Balkan, 2006; Sarazin et al., 2007). In addition, cued recall based on category cues makes it easier to discriminate between patients with AD, patients with amnesic mild cognitive impairment (aMCI), and HCs because, when compared to free recall, cued recall based on category cues improves retrieval of test material more for aMCI patients than AD patients and HCs show a ceiling effect (Grober & Buschke, 1987; Grober et al., 1988; Saka et al., 2006;

Sarazin et al., 2007). A limitation of cued recall based on category cues, however, is that retrieval of test material is primarily improved in the early stages of AD and to a lesser degree during the progression of the disease (Tounsi et al., 1999). Retrieval of test material might be further improved by expanding encoding specificity beyond category cues.

The Visual Association Test (VAT; Lindeboom & Schmand, 2003) is an episodic memory test that expands encoding specificity by presenting to the patient unexpected visual associations of a target stimulus and a cue interacting with one another, such as a rabbit holding a kite. Without being told that the associations are to be remembered, the patient is asked to, first, name the cue and, then, name the cue interacting with the target. Within this incidental learning paradigm, the unexpected visual associations induce elaboration of the stimuli, which establishes stronger and longer lasting memory traces (Craik & Lockhart, 1972; Tulving & Madigan, 1970). This is supported by research showing that older persons do better on this type of cued recall task, compared to a cued recall task in which the same material is presented side-by-side as paired associates, and which is based on intentional learning (Lindeboom, Schmand, Tulner, Walstra & Jonker, 2002). This finding is quite remarkable, because intuitively one would expect the conditions for cued recall to be more favourable for intentional learning. But, conversely, through the incidental learning paradigm and the interactive presentation, the visual associations are elaborated on to such an extent that it exceeds the advantage of intentional learning. The VAT has been used in various studies, showing a high validity in predicting progression to AD and the ability to differentiate between HCs, aMCI patients and AD patients (Dierckx et al., 2009; Fuchs, Wiese, Altiner, Wollny & Pentzek, 2011; Lindeboom et al., 2002; Lindeboom & Schmand, 2003; Reesink et al., 2010; Vlies, van der et al., 2009). To enable repeated assessment of the episodic memory impairment in AD parallel versions of the VAT were recently constructed (Lindeboom, Schmand, Meyer & Jonghe, de, 2014).

The purpose of this paper was to examine the potential use of the VAT for monitoring episodic memory impairment in patients with aMCI or AD. If a test is unable to monitor episodic memory impairment within these groups a floor effect is expected, i.e. test performances will level off as global cognitive impairment increases. In a regression model, this would be evident in a curvilinear trend in which the curve flattens out at the low end of the measurement range, corresponding with a quadratic model. To investigate this within and between the groups HCs, aMCI and AD, we used hierarchical multiple regression analyses to determine if the relationship between global cognitive impairment, as measured with the Cognitive Screening Test (CST; Graaf & Deelman, 1991), and test performance on the VAT or RAVLT could be explained by a linear or quadratic model. If the quadratic trend significantly increased predictive capacity after the linear trend had been taken into account, this was indicative of a floor effect. We hypothesized that expanded encoding specificity, as implied in the VAT, will show a smaller floor effect in aMCI and mild AD patients than word list learning.

Methods

Participants

Our sample contained 35 patients with AD, 39 patients with aMCI, and 51 HCs. The demographic characteristics of these groups are listed in Table 1. We scored education according to the Dutch scoring system Verhage (Verhage, 1964), which ranges from one to seven, with one being less than elementary education and seven being university. The Verhage system overlaps with the International Standard Classification of Education 1997 (ISCED 1997) scales (UNESCO, 2006). We recruited the patients from the Medical Center Alkmaar geriatric outpatient memory clinic and the HCs from the informal caregivers of these patients, all of whom were close family members or spouses of the patients. Participants were

eligible for inclusion if they were 65 years or older, and had intact or corrected to normal hearing and vision. Patients were eligible for inclusion if they met the consensus diagnostic criteria for aMCI or AD (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision [DSM-IV-TR]; American Psychiatric Association, 2000; Dubois et al., 2007; Petersen et al., 2001). The diagnosis of aMCI or AD was discussed in a multidisciplinary team of five geriatricians, a neurologist, a clinical neuropsychologist, and an old-age psychiatrist. Routine diagnostic procedures were based on a comprehensive geriatric assessment including physical examination, laboratory tests, MRI or CT imaging, and standardized neuropsychological evaluation. HCs were included if they had a Cognitive Screening Test (CST; Graaf & Deelman, 1991) score of 13 or higher. Exclusion criteria for all participants were severe brain damage, brain tumour, epilepsy, multiple sclerosis, Parkinson's disease, psychiatric disorder, delirium, or treatment in the past for addiction to alcohol or drugs. All participants provided written informed consent. The study was approved by the local ethical review board.

Table 1. Participants demographic characteristics and test performances according to group

| | HC (<i>n</i> = 51) | aMCI (<i>n</i> = 39) | AD (<i>n</i> = 35) |
|-----------|---------------------|-----------------------|---------------------|
| Gender | 64.7 | 59.0 | 40.0 |
| Age | 74.1 (5.7) | 78.5 (6.6) | 79.7 (6.3) |
| Education | 4.8 (1.3) | 4.2 (1.4) | 4.5 (1.4) |
| VAT | 10.8 (1.6; 5-12) | 6.7 (3.9; 0-12) | 3.9 (4.0; 0-11) |
| RAVLT | 38.7 (10.8; 19-59) | 21.9 (6.8; 8-34) | 17.2 (7.5; 5-38) |
| CST | 18.7 (1.4; 14.5-20) | 15.7 (3.0; 7-20) | 13.2 (3.9; 5-19) |
| MMSE | 28.0 (1.1; 24-29) | 25.1 (3.3; 14-29) | 22.0 (5.2; 11-28) |

Notes: HC = healthy elderly control; aMCI = amnesic mild cognitive impairment; AD = Alzheimer's Disease; number of subjects (i.e., *n*); gender (proportion of females); age in years (mean and standard deviation); education is scored according to the Verhage system, with 1 = less than elementary education, and 7 = university (mean and standard deviation); VAT = Visual Association Test (mean, standard deviation and range of scores); RAVLT = Rey Auditory Verbal Learning Test (mean, standard deviation and range of scores); CST = Cognitive Screening Test (mean, standard deviation and range of scores); MMSE = Mini-Mental State Examination (mean, standard deviation and range of scores).

Design and Measurements

This is a cross-sectional study examining VAT and RAVLT lower range test scores across groups of HCs, and patients with aMCI or AD compared to global cognitive impairment as measured with the CST. Tests were administered to patients as part of the standardized neuropsychological evaluation and at the same time, but separately, to HCs.

The VAT (Lindeboom & Schmand, 2003) consists of six items of black and white pictures (line drawings). A target picture to be remembered is linked to a cue picture by associating these through an interaction, e.g. a rabbit holding a kite (See Figure 1). The subject is not told that the task is a memory test (incidental learning). First, the subject is instructed to name the six cues (step one). Next, the cues are shown in interaction with the corresponding targets, which have to be named as well (step two). After the first two steps, all six cues are shown again one by one, and the subject is instructed to recall the associated target (step three). Finally, steps two and three are repeated. VAT total scores range from 0 to 12 points. The VAT has four parallel versions, A, B, C and D, two of which we used: versions C and D. These have high reliability, with a parallel form reliability of .91 for both versions (Lindeboom et al., 2014). Participants received either version C or D at random.

The Dutch version of the RAVLT was used (15WT; Saan & Deelman, 1986). It presents a list of 15 unrelated words that are read out loud by the test leader in five repeated trials. Before presentation, the subject is instructed to remember all the words, thus promoting intentional learning. Immediately after each presentation, the subject is instructed to recall as many words as possible. Thus, learning was assessed by the total number of words recalled over five trials. The possible score range was 0 - 75 points.

The CST (Graaf & Deelman, 1991) is a measure of severity of global cognitive impairment. It consists of 20 questions relating to orientation in time, place and person, and elementary factual knowledge. The score ranges from 0 to 20 points, with high scores

indicating lack of impairment. The CST was developed based on the short portable mental status questionnaire (SPMSQ; Pfeiffer, 1975) and the mental status questionnaire (MSQ; Kahn, Goldfarb, Pollack & Peck, 1960). Validity has been demonstrated as it highly correlates with the Mini-Mental State Examination (MMSE; Folstein, Folstein & HcHugh, 1975), $r = 0.86$, (Dautzenberg, Schmand, Vriens, Deelman & Hooijer, 1991) and, in a large-scale sample study, $n = 4051$, it was equally predictive of dementia as the MMSE, with an area under the curve of .911 (Schmand, Deelman, Hooijer, Jonker & Lindeboom, 1996). From that research mean MMSE scores were established as a function of CST scores, which enabled us to estimate MMSE scores. The CST has a high reliability, with an internal consistency (Cronbach's alpha) of 0.89 (Dautzenberg et al., 1991) and a test-retest reliability of 0.80 (Schmand et al., 1996).



Figure 1. Sample of cue (left) and interaction with target (right).

Data Analysis

Preliminary analyses showed the parametric assumptions were violated for age in the aMCI group, for education and VAT in all groups, and for RAVLT and CST in the HCs group. When parametric assumptions were violated we used nonparametric tests. We compared

mean VAT, RAVLT and CST total scores and demographic characteristics between groups of HCs, aMCI patients and AD patients using Kruskal-Wallis tests, Mann-Whitney tests, Pearson's chi-square test and t-tests, as appropriate. To investigate floor effects, VAT and RAVLT total scores were transformed into z-scores and plotted against the range of global cognitive impairment as measured with CST total scores, and hierarchical multiple regression analyses were used to examine whether a linear or quadratic model fitted the data best. The CST total score served as a predictor of test performance in the regression model, with CST total score representing a linear trend, and the squared CST total score representing a quadratic trend. We first entered the CST total score into the regression model and then entered the squared CST total score. These predictors were maintained within the model only if they significantly increased the percentage of explained variance. This resulted in either a linear model as best fit for the data if only the linear trend significantly increased the percentage of explained variance, or a quadratic model as best fit for the data if the linear trend together with the quadratic trend both significantly increased the percentage of explained variance. In our study, the a priori criterion for establishing the presence of a floor effect was if a quadratic model best fitted the data, the quadratic model indicating a floor effect and the total scores below the constant in this model indicating the range of the floor effect. Preliminary analyses were conducted to investigate the parametric assumptions of multiple regression, which showed that the assumption of multicollinearity was violated. However, this was as expected, because the CST total score and squared CST total score are based on the same variable. Also, as expected, the assumption of linearity was violated for the quadratic model. If demographic variables showed significant differences between groups, we investigated the effect of such covariates on the hierarchical multiple regression models.

Results

Table 1 shows demographic characteristics and test performances of the groups. Education ($H(2) = 3.47, p = .176$) and gender ($\chi^2(2) = 5.34, p = .069$) did not differ between groups, however, age did differ between groups ($H(2) = 17.75, p < .001$). Mean VAT and RAVLT total scores differentiated HCs from aMCI patients (VAT: $U = 292.5, p < .001$; RAVLT: $U = 201.5, p < .001$) and aMCI patients from AD patients (VAT: $U = 415.5, p < .01$; RAVLT: $t(72) = 2.80, p < .01$). Mean CST total scores differentiated HCs from aMCI patients ($U = 346.0, p < .001$) and aMCI patients from AD patients ($t(72) = 3.10, p < .01$).

Table 2 shows VAT test performances of HCs, aMCI patients, and AD patients in relation to severity of global cognitive impairment. CST total scores explained 47.2% of the variance in VAT test performances ($F(1,123) = 110.05, p < .001$). Addition of the squared CST total scores did not significantly increase the percentage of explained variance (R squared change = .005, F change (1,122) = 1.19, $p = .277$). This effect remained after controlling for the effect of age (F change (1,121) = 1.08, $p = .301$). Thus, a linear regression model best fitted the VAT data, indicating that the VAT did not show a floor effect (See Figure 2).

Table 2. Summary of Hierarchical Regression Analysis for CST Variables Predicting VAT Total Scores ($N = 125$)

| Variable | <i>B</i> | <i>SE B</i> | β |
|--------------------------|----------|-------------|---------|
| Step 1 (linear model) | | | |
| Constant | -5.65 | 1.29 | |
| CST | 0.81 | 0.07 | .68* |
| Step 2 (quadratic model) | | | |
| Constant | -1.70 | 3.84 | |
| CST | 0.22 | 0.55 | .18 |
| CST squared | 0.02 | 0.01 | .50 |

Note. $R^2 = .472$ for Step 1; $\Delta R^2 = .005$ for Step 2 ($p = .277$). * $p < .001$. CST = Cognitive Screening Test; VAT = Visual Association Test.

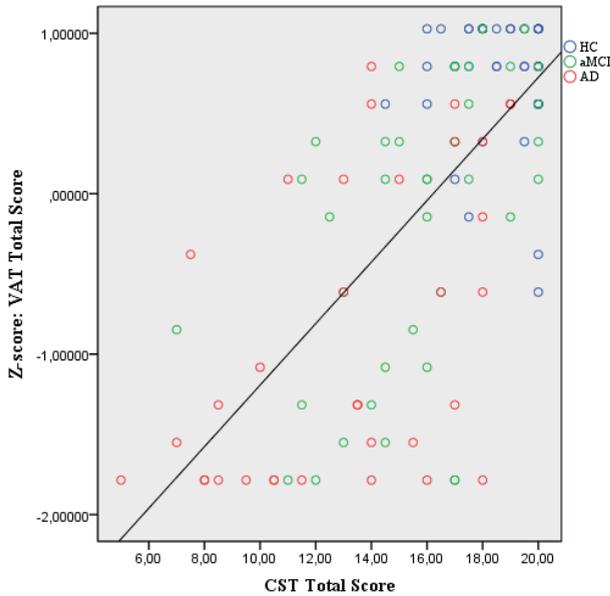


Figure 2. Z-scores VAT total scores compared to CST total scores for AD patients ($n = 35$), aMCI patients ($n = 39$) and HC participants ($n = 51$).

Notes: AD = Alzheimer's Disease; aMCI = amnestic mild cognitive impairment; HC = healthy elderly control; VAT = Visual Association Test; CST = Cognitive Screening Test.

Table 3 shows RAVLT test performances of HCs, aMCI patients, and AD patients in relation to severity of global cognitive impairment. CST total scores explained 43.3% of the variance in RAVLT test performances. Addition of the squared CST total scores significantly increased the percentage of explained variance to 49.1% ($F(2,122) = 58.93, p < .001, R$ squared change = .058, F change (1,122) = 13.96, $p = < .001$). This effect remained after controlling for the effect of age (F change (1,121) = 13.57, $p < .001$). Thus, a quadratic regression model best fitted the RAVLT data, indicating a floor effect for the RAVLT (See Figure 3). The range of which was total scores of 29 (5,8 per trial) or below, which was scored by 90% of aMCI patients and 94% of AD patients.

Table 3. Summary of Hierarchical Regression Analysis for CST Variables Predicting RAVLT Total Scores

(*N* = 125)

| Variable | <i>B</i> | <i>SE B</i> | β |
|--------------------------|----------|-------------|---------|
| Step 1 (linear model) | | | |
| Constant | -11.13 | 4.07 | |
| CST | 2.37 | 0.24 | .65** |
| Step 2 (quadratic model) | | | |
| Constant | 29.34 | 11.50 | |
| CST | -3.73 | 1.65 | -1.03* |
| CST squared | 0.21 | 0.05 | 1.70** |

Note. $R^2 = .433$ for Step 1; $\Delta R^2 = .058$ for Step 2 ($p < .001$). * $p < .05$. ** $p < .001$. CST = Cognitive Screening Test; RAVLT = Rey Auditory

Verbal Learning Test.

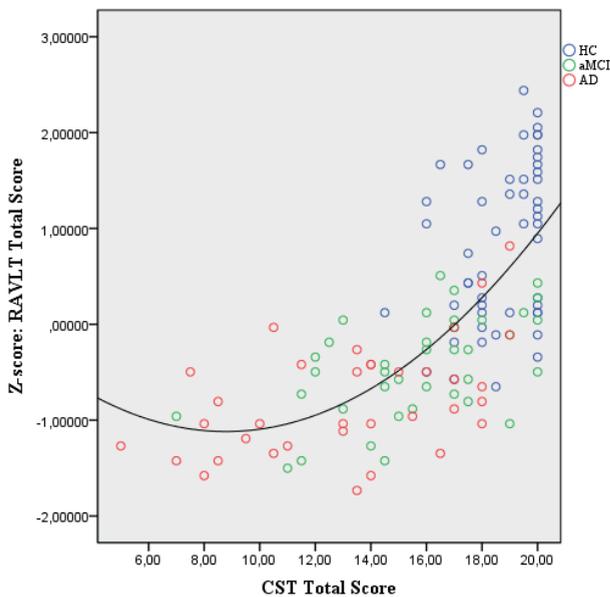


Figure 3. Z-scores RAVLT total scores compared to CST total scores for AD patients ($n = 35$), aMCI patients ($n = 39$) and HC participants ($n = 51$).

Notes: AD = Alzheimer’s Disease; aMCI = amnestic mild cognitive impairment; HC = healthy elderly control; RAVLT = Rey Auditory Verbal Learning Test; CST = Cognitive Screening Test.

Discussion

This study highlights the importance of examining floor effects of cognitive tests that are used to measure deterioration in patients with mild cognitive impairment and mild dementia.

Results showed a floor effect for the RAVLT, a well-known auditory verbal learning test, with 90% of all patients in this sample scoring in the lower range. No floor effect was found in healthy elderly controls. By contrast, the VAT measured episodic memory decline in a linear fashion in both controls and patients across the range of global cognitive impairment. Thus, the VAT is potentially useful for monitoring episodic memory decline beyond the distinction between normal and abnormal memory function.

These findings may have implications for clinical trials and for monitoring memory clinic outpatients, where sensitive measures of change are needed to evaluate intervention outcomes and, similarly, to assess conversion from aMCI to mild AD. The present findings suggest that a lengthy word list memory task may not be the first test of choice for monitoring episodic memory decline, once a diagnosis of aMCI or AD has been established.

Our results are in line with those in previous studies. The general picture that emerges from studies using the VAT or the RAVLT in HCs, aMCI and AD samples shows that average test scores discriminate between groups with high sensitivity and specificity (Dierckx et al., 2009; Eckerström et al., 2013; Fuchs et al., 2011; Petersen et al., 1999). Contrary to previous studies, however, we investigated memory test scores in relation to global cognitive impairment, thus introducing a third variable, which enabled us to examine whether episodic memory tests can differentiate between patients with mild to more severe global cognitive decline. We showed that, although the RAVLT statistically differentiated aMCI and AD, this test is of less value for monitoring purposes.

By contrast, the VAT has the potential to monitor AD patients' episodic memory impairment, due to the encoding specificity it provides, which improves retrieval by means of

cued recall based on incidental learning of cues and targets that interact with one another and do so in an unexpected way. This interaction, by and large, is based on the classical peg word mnemonic, in which information to be remembered must be linked to a known sequence of pegs by forming a mental image of two interacting objects. The VAT builds on this paradigm by providing the mental image by means of the visual association and indicating the peg by means of the cue for retrieval (Lindeboom & Schmand, 2003). These visual associations are unexpected and, thus, within the incidental learning paradigm, are elaborated on. This strongly links cue and target and, as such, stronger and longer lasting memory traces are established (Craik & Lockhart, 1972; Tulving & Madigan, 1970). In addition, encoding is less burdensome for patients, because the VAT is a short test that is easy to administer and that promotes incidental learning, which makes it especially suitable for vulnerable people such as patients with AD.

Strong points of our study are the complete data set, the inclusion of HCs and patients with aMCI or mild AD, and the use of a regression-based method for examining floor effects. By using hierarchical multiple regression analysis, we were able to determine if the quadratic trend of a measure of global cognitive impairment, i.e. a floor effect, significantly increased the percentage of explained variance in addition to the linear trend. We investigated this in a large sample that covered the whole continuum of the disease, which is why we consider our findings to be robust.

A limitation that needs to be discussed is the ecological validity of the VAT. It could be argued that the VAT aids retrieval too much, thus not being applicable to daily life. However, in daily life, like in the VAT, knowledge is often linked with the environment, which constantly offers cues to memory. For example, knowledge of how to operate a microwave oven is recalled by looking at the control panel of the microwave itself. Also, a lot of

knowledge in daily life is acquired, like the VAT, outside a formal learning setting and by incidental learning (Marsick & Watkins, 2001).

A second limitation is the use of a screening test, i.e. the CST, to determine the severity of global cognitive impairment, which might be less sensitive to mild cognitive deficits like aMCI. This could have had an effect on our results, by a restriction of range for these groups of patients. However, as can be seen in table 1, standard deviations and, thus, variances of CST total scores increased substantially from HCs to aMCI to AD, with an increase of more than 300% from HCs to aMCI and an increase of more than 70% from aMCI to AD. In view of these results, we consider the CST to be a sufficiently sensitive measure of global cognitive impairment in these patients.

A third limitation is the fact that we did not measure episodic memory changes longitudinally. However, our study did include HCs and patients with aMCI or AD, and memory assessment was not only focused on group membership, but also on global cognitive impairment, thus enabling measurement of differences within groups. Also, in various studies, the VAT has shown high validity as a diagnostic tool for AD as well as the ability to differentiate between HCs, aMCI patients and AD patients (Dierckx et al., 2009; Fuchs et al., 2011; Lindeboom et al., 2002; Lindeboom & Schmand, 2003; Reesink et al., 2010; Vlies, van der et al., 2009). Although this strengthens the generalizability of our findings, future research should test episodic memory decline using a longitudinal design, in which HCs and patients from our study are retested. By doing so, the present findings can be integrated into an analysis of longitudinal change in memory function as a function of the nature of support of encoding and retrieval.

To accurately monitor episodic memory impairment in aMCI patients and mild AD patients, we need to have cognitive measures that measure change. As our findings show, word list learning tests can be too difficult for patients with aMCI and AD, whereas the visual

association paradigm of the VAT can differentiate between patients with mild to more severe global cognitive decline. Ideally, these two tests are used in conjunction with each other, since word list learning tests are especially sensitive for the diagnosis of aMCI and AD, while the VAT is potentially valuable for monitoring disease progression. As many AD patients are assessed more than once after the initial diagnosis, it is essential to choose a sensitive measure for monitoring the disease. The VAT has the potential to do so.

Acknowledgments

We would like to thank the Department of Geriatric Medicine of the Medical Center Alkmaar for the possibility to do our research.

Disclosure of Interest

S.R.A.M and J.F.M.J are the authors of versions C and D of the Visual Association Test.

Funding

This work was supported by the Foreest Medical School Alkmaar [grant number FIO1308].

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: Author.
- Bigler, E. D., Rosa, L., Schultz, R., Hall, S., & Harris, J. (1989). Rey-auditory verbal learning and rey-osterrieth complex figure design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology*, *45*, 277-280.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and*

assessment. *An introduction to tests and measurement* (8th ed., p.p. 256, 259). New York: McGraw-Hill.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.

Dautzenberg, P. L. J., Schmand, B., Vriens, M. T. S., Deelman, B. G., & Hooijer, C. (1991). De validiteit van de cognitieve screening-test en de 'mini-mental state examination' bij een groep oudere ziekenhuispatiënten [Study of the validity of the cognitive screening test and the mini-mental state examination in an older hospital population]. *Nederlands Tijdschrift voor Geneeskunde*, *135*, 850-855.

Dierckx, E., Engelborghs, S., Raedt, R. de, Buggenhout, M. van, Deyn, P. P. de, Verté, D., et al. (2009). Verbal cued recall as a predictor of conversion to Alzheimer's disease in mild cognitive impairment. *International Journal of Geriatric Psychiatry*, *24*, 1094-1100.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, *6*, 734-746.

Eckerström, C., Olsson, E., Bjerke, M., Malmgren, H., Edman, A., Wallin, A., et al. (2013). A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts conversion from mild cognitive impairment to dementia. *Journal of Alzheimer's Disease*, *36*, 421-431.

Folstein, M. F., Folstein, S. F., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.

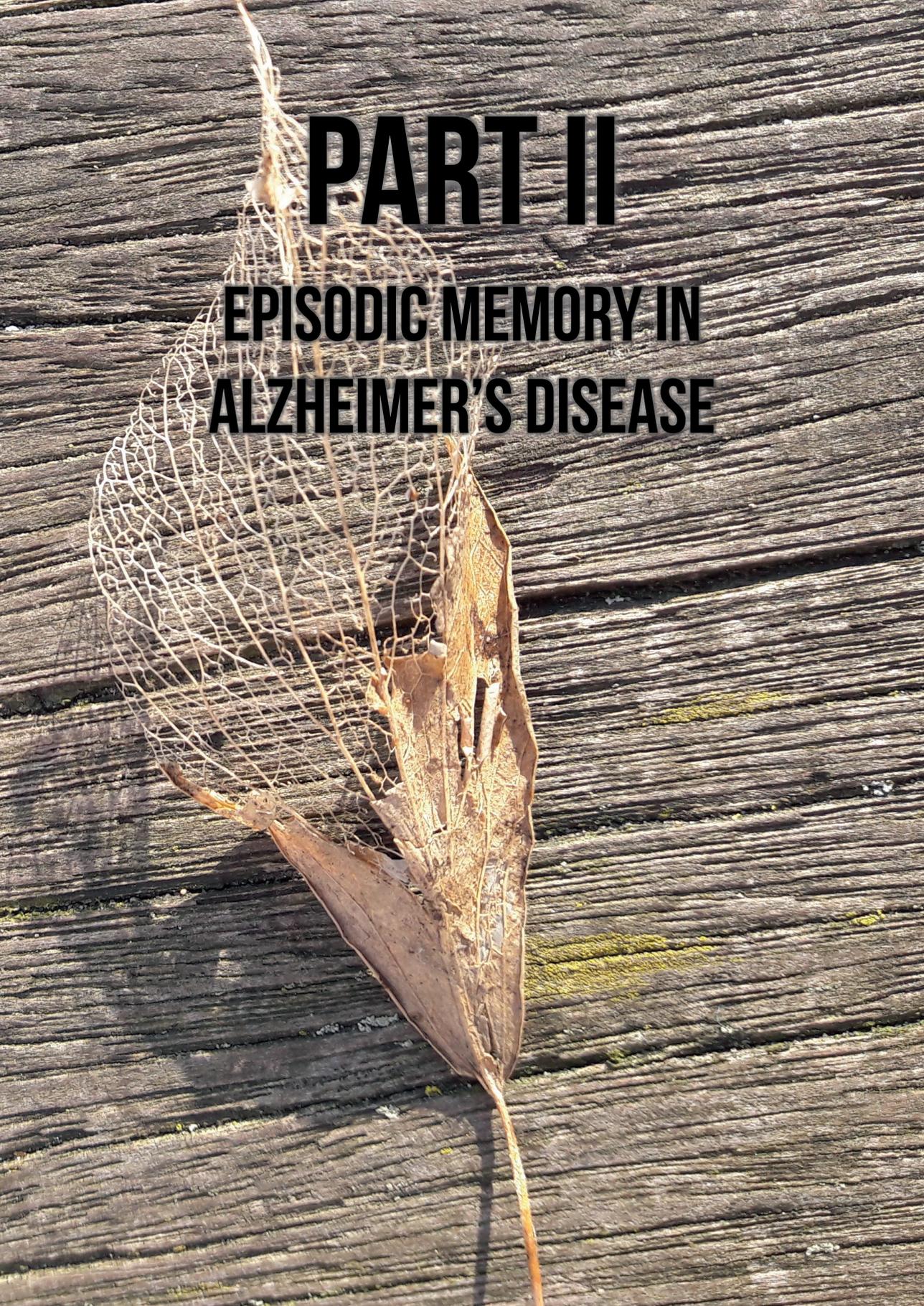
Fuchs, A., Wiese, B., Altiner, A., Wollny, A., & Penzek, M. (2011). Cued recall and other cognitive tasks to facilitate dementia recognition in primary care. *Journal of*

- American Geriatric Society*, 60, 130-135.
- Graaf, A. de, & Deelman, B.G. (1991). *Cognitive Screening Test. Manual for administration and scoring*. Lisse: Swets en Zeitlinger bv.
- Grober, B., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental neuropsychology*, 3, 13-36.
- Grober, E., Buscke, H., Crystal, H., Bang, S., & Dresner, R. N. C. (1988). Screening for dementia by memory testing. *Neurology*, 38, 900-903.
- Kahn, R. L., Goldfarb, A. I., Pollack, M., & Peck, A. (1960). Brief objective measures for the determination of mental status in the aged. *The American Journal of Psychiatry*, 117, 326-328.
- Kessels, R. P. C., Rijken, S., Joosten-Weyn Banningh, L. W. A., Schuylenborgh-van Es, N. van & Olde Rikkert, M. G. M. (2009). Categorical spatial memory in patients with mild cognitive impairment and alzheimer dementia: Positional versus object-location recall. *Journal of the International Neuropsychological Society*, 16, 200-204.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test. Manual*. Leiden: PITS bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & Jonghe, J. F. M. de (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73, 126-133.
- Marsick, V. J., & Watkins, K. E. (2001). Informal and incidental learning. *New Directions for Adult and Continuing Education*, 89, 25-34.
- Pasquier, F., Grymonprez, L., Lebert, F., & Linden, M. van der (2001). Memory impairment differs in frontotemporal dementia and Alzheimer's disease. *Neurocase*, 7, 161-171.
- Petersen, R. C., Doody, R., Kurz, A. Mohs, R. C., Morris, J.C., Rabins, P. V., et al. (2001).

- Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985-1992.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment. *Archives of Neurology*, 56, 303-308.
- Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society*, 23, 433-441.
- Pillon, B., Deweer, B., Michon, A., Malapani, C., Agid, Y., & Dubois, B. (1994). Are explicit memory disorders of progressive supranuclear palsy related to damage to striatofrontal circuits? Comparison with Alzheimer's, Parkinson's and Huntington's diseases. *Neurology*, 44, 1264-1270.
- Reesink, F. E., Lemstra, A. W., Dijk, K. D. van, Berendse, H. W., Berg, W. D. J. Klein, M. van de, et al. (2010). CSF α -synuclein does not discriminate dementia with Lewy bodies from Alzheimer's disease. *Journal of Alzheimer's Disease*, 22, 87-95.
- Rey, A. (1964). *L'examen Clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Saan, R. J., & Deelman, B. G. (1986). *15-Woorden Test A en B* [15-Word Test A and B]. In Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (Eds.) (2012). *Handboek neuropsychologische diagnostiek* [Handbook neuropsychological assessment] (2nd ed., p.p. 267-282). Amsterdam: Pearson Assessment and Information B.V.
- Saka, E., Mihci, E., Akif Topcuoglu, M., & Balkan, S. (2006). Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Archives of Clinical Neuropsychology*, 21, 745-751.
- Sarazin, M., Berr, C., Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., et al. (2007). Amnesic syndrome of the medial temporal type identifies prodromal AD: A

- longitudinal study. *Neurology*, 69, 1859-1867.
- Schmand, B., Deelman, B. G., Hooijer, Ch., Jonker, C., & Lindeboom, J. (1996). De itemreeks van de cognitieve screening test vergeleken met die van de mini-mental state examination [The items of the cognitive screening test compared to those of the mini-mental state examination]. *Tijdschrift voor Gerontologie en Geriatrie*, 27, 29-33.
- Schmidt, M. (1996). *Rey Auditory Verbal Learning Test. A handbook*. Los Angeles: Western Psychological Services.
- Tounsi, H., Deweer, B., Ergis, A., Linden, M. van der, Pillon, B., Michon, A., et al. (1999). Sensitivity to semantic cueing: An index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 13, 38-46.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21, 437-484.
- UNESCO (2006). *International Standard Classification of Education: ISCED 1997 (re-edition)*. Montreal: UNESCO-UIS.
- Verhage, F. (1964). Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar [Intelligence and age: Research within Dutch participants aged from 12 to 77 years old]. Dissertation. Assen: Van Gorcum.
- Vlies, A. E. van der, Verwey, N. A., Bouwman, F. H., Blankenstein, M. A., Klein, M., Scheltens, P., et al. (2009). CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology*, 72, 1056-1061.
- Woodward, J. L., Dunlosky, J., & Salthouse, T. A. (1999). Task decomposition analysis of intertrial free recall performance on the rey auditory verbal learning test in normal aging and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 21, 666-676.





PART II
EPISODIC MEMORY IN
ALZHEIMER'S DISEASE

3

Chapter 3

**EPISODIC RECOGNITION MEMORY
BASED ON INCIDENTAL LEARNING OF
VISUAL ASSOCIATIONS IS LARGELY
PRESERVED COMPARED TO RECALL
IN AMNESTIC MILD COGNITIVE
IMPAIRMENT AND MILD ALZHEIMER'S
DISEASE**

Sascha R.A. Meyer, Leo Boelaarts, Jaap Lindeboom, Jos F.M. de Jonghe, and
Rudolf W.H.M. Ponds

Applied Neuropsychology: Adult, 2019. Advanced online publication.
doi:10.1080/23279095.2019.1703705

Abstract

We investigated preserved episodic recognition memory based on incidental learning of visual associations in Alzheimer's disease (AD). In a cross-sectional design, we analyzed episodic memory score profiles of patients with amnesic mild cognitive impairment (a-MCI) ($n = 42$) or mild AD ($n = 19$) who had hippocampal atrophy, and healthy elderly controls ($n = 43$). The Visual Association Test-Extended served as a measure of episodic memory. Multiple-choice cued recognition was compared with paired associate recall and free recall within groups. Results showed that patients recognized learned material much better compared to when they had to recall material, resulting in large effect sizes (Cohen's d) ranging from 1.3 to 3.5. We conclude that episodic recognition memory based on incidental learning of visual associations is largely preserved when compared to recall in a-MCI and mild AD. This suggests that the episodic memory impairment in AD may be characterized as a retrieval impairment rather than a consolidation impairment, indicating that preserved recognition compared to recall may be compatible with AD being the correct diagnosis. Measuring the episodic memory impairment in AD may benefit from using tests that capture different aspects of memory processes such as incidental learning of visual associations.

Introduction

Alzheimer's disease (AD) is prevalent in older persons; affecting 60 to 70% of all 50 million dementia cases worldwide (World Health Organization, 2017). The cognitive hallmark of AD in neuropsychological testing is an encoding and consolidation deficit of newly learned information. Current core diagnostic criteria for probable AD state that the episodic memory impairment consists of memory performance that does not improve with recognition testing after a patient showed a deficit on recall, which is considered to be clinically helpful in differentiating AD from other dementias (Dubois et al., 2007). However, AD might not just be characterized by impaired consolidation, as recent findings suggest that recognition memory might be preserved in AD (Roy et al., 2016).

Research suggests that in general enhanced engagement with test material establishes more elaborate, longer lasting, and stronger memory traces (Craik & Lockhart, 1972). This suggests that also in case of AD learned information might be remembered better when employing enriched learning. Furthermore, research suggests that in general the memory performance of subjects improves after recall if they are subsequently tested by means of cueing or recognition (Tulving & Madigan, 1970). Indeed, although episodic recall memory is impaired in patients with amnesic mild cognitive impairment (a-MCI) or AD (Gifford et al., 2015; Kessels, Rijken, Joosten-Weyn Banningh, van Schuylenborgh-van Es, & Olde Rikkert, 2009; Meyer et al., 2016), a meta-analysis of 47 studies showed that episodic recognition memory was significantly better than recall in incident AD (Bäckman, Jones, Berger, Laukka, & Small, 2005).

Research suggests that differences in the underlying neural processes may explain why recognition memory might be better than recall in AD patients. Neural findings show that recall is strongly associated with hippocampal volumes in AD (Deweert et al., 1995), and that, in patients with brain injury, the hippocampus was activated by recall but not by recognition

(Skinner & Fernandes, 2007). Also, in healthy persons, recognition depends on additional cortices, as it is associated with increased activity in frontal, cingulate and parietal cortices, and the thalamus (Montaldi, Spencer, Roberts, & Mayers, 2006). Interestingly, healthy elderly controls and a-MCI patients who were trained to recognize a salient feature in a picture and link this with an object in a mental image, also showed increased activation in frontal and parietal cortices (Hampstead, Stringer, Stilla, & Sathian, 2019).

Most studies that showed improvement of memory performance on recognition testing after recall in AD used verbal memory tests (Bäckman et al., 2005). However, memory performance may improve even more by means of visual memory tests because of the so-called picture superiority effect, i.e. pictures are more easy to recall than words (Shepard, 1967), which applies to normal subjects for free recall (Paivio, Rogers, & Smythe, 1968), associative recognition (Hockley, 2008), and recognition (Shepard, 1967), and also to a-MCI patients and AD patients for recognition (Alley, 2012).

The Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017) is an episodic memory test that uses pictures and enhances the engagement with test material by means of visual associative learning, i.e. learning of pictures that depict an unexpected interaction between a target stimulus and a stimulus cue, such as a dog riding a bicycle (See Figure 1). This interaction is based on a classical mnemonic in which a target stimulus has to be linked to a cue by forming a mental image (Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002). These visual associations are learned within an incidental learning paradigm. A subject learns incidentally when he is unaware that he is learning. The VAT-E consists of three measures of memory, i.e. paired associate recall, free recall, and multiple-choice cued recognition.

The purpose of our study was to improve the assessment of memory dysfunction in AD. Therefore, we investigate preserved episodic recognition memory using incidental

learning of visual associations in patients with a-MCI or mild AD. Since the VAT-E combines recall and recognition trials in one single test, it will enable clinicians to see at a glance if a patients' memory performance improves with recognition testing or not. We hypothesized that test results of a-MCI and mild AD patients would significantly improve on VAT-E multiple-choice cued recognition compared to VAT-E paired associate recall and VAT-E free recall.

Contrary to our previous research on these measures (Meyer, de Jonghe, Schmand, & Ponds, 2019), we now focus on analyses of memory improvement within groups. This was not possible in our earlier paper, as we analyzed each measure separately in relation to global cognitive impairment. Also, contrary to our previous research (Meyer et al., 2019), we included a-MCI or AD patients not only based on clinical symptoms, but also based on medial temporal lobe atrophy supportive for the diagnosis of probable AD as evidenced by magnetic resonance imaging (Scheltens et al., 1992).

Methods

Participants

A total of 43 healthy elderly controls, 42 patients with a-MCI patients (32 amnesic MCI single domain and 10 amnesic MCI multiple domain), and 19 patients with mild AD were included. Patients were recruited amongst consecutively referred geriatric outpatients of the memory clinic of the Northwest Medical Center Alkmaar, a large research and teaching hospital in the Netherlands. We recruited healthy elderly controls amongst the informal caregivers of these patients, i.e. their spouses or family members. Written informed consent was provided by all participants. The study was approved by the local medical ethical review board.

Participants were eligible for inclusion provided they were 65 years or older and visual and auditory acuity were intact or corrected to normal. We excluded participants if they had severe traumatic brain damage or brain disease (e.g. brain tumor, epilepsy, multiple sclerosis, or Parkinson's disease), a severe psychiatric disorder, delirium or a history of alcohol or drug abuse. Healthy elderly controls were excluded if they had a score of less than 13 on the Cognitive Screening Test (De Graaf & Deelman, 1991), and if they scored at percentile 5 or below, corrected for gender, age, and education, on delayed recall of the Dutch version of the Rey Auditory Verbal Learning Test (Saan & Deelman, 1986). The healthy elderly controls group was the same as in our previous research (Meyer et al., 2019); we excluded one case due to incorrect age; intermediate analyses showed this did not affect our previous findings (Meyer et al., 2019). The a-MCI and AD groups were subsets of the groups in our previous research (Meyer et al., 2019); we now included patients not only based on clinical symptoms, but also based on biomarker evidence. Therefore, patients were included if magnetic resonance imaging findings showed evidence of AD as indicated by medial temporal lobe atrophy grade two or higher (Scheltens et al., 1992). In accordance with the research of Scheltens et al., (1992) the medial temporal lobe atrophy score was assessed visually by a multidisciplinary team (geriatric consultants, a neurologist, a clinical neuropsychologist, and an old-age psychiatrist), and set as a definite score after consensus was reached.

All patients were assessed using a standard neuropsychological test protocol that, besides several tests and questionnaires concerning the cognitive domains under discussion, included the study's outcome measure VAT-E. Notably, VAT-E scores were not used during the diagnostic process. To ensure this, the VAT-E score sheets were removed from patients' hospital files, thus blinding the clinical neuropsychologist to the VAT-E scores. Healthy elderly controls took the VAT-E and the measures used for exclusion, i.e. the Cognitive Screening Test and the Rey Auditory Verbal Learning Test. To prevent contamination

between healthy elderly controls and patients due to them sharing stories about the test protocol or the test items, we had the healthy elderly controls take the tests at the same time but in a separate room as where the patients were examined.

Diagnoses of a-MCI or AD were based on internationally accepted criteria (Dubois et al., 2007; Petersen et al., 2001; Petersen, 2004). Patients underwent a routine diagnostic program that consisted of a comprehensive geriatric assessment including laboratory tests and neuroimaging, i.e. magnetic resonance imaging, and a standard neuropsychological examination. A final diagnosis of a-MCI or AD was made after all clinical data had been discussed by members of a multidisciplinary team, which included geriatric consultants, a neurologist, a clinical neuropsychologist, and an old-age psychiatrist.

Measures

The Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017) is an episodic memory test that provides several measures of memory, i.e. cued recall, free recall and recognition. Development of the VAT-E builds on research into the Visual Association Test (VAT; Lindeboom, Schmand, Meyer, & de Jonghe, 2014). Various studies have shown that the VAT is a highly valid measure that can be used to predicting AD development and that can differentiate between healthy elderly controls, a-MCI, and AD (Dierckx et al., 2009; Fuchs, Wiese, Altiner, Wollny, & Penzek, 2012; Lindeboom, Schmand, Meyer, & de Jonghe, 2014; Lindeboom et al., 2002; Meyer, Spaan, Boelaarts, Ponds, Schmand, & de Jonghe 2016; Reesink et al., 2010; Van der Vlies et al., 2009). In line with previous work on the VAT, sound psychometric qualities were established for the VAT-E in studies that used it to differentiate between healthy elderly controls, a-MCI, and AD, and correlated it with delayed free recall of the Rey Auditory Verbal Learning Test ($r = .69$ to $.85$) (Meyer & de Jonghe, 2017; Meyer, de Jonghe, Schmand, & Ponds, 2019).

All VAT-E trials were administered using the same instruction and the same order of item presentation in accordance with the standardized test instruction as stated in the manual of the test (Meyer & de Jonghe, 2017). The VAT-E consists of black-and-white line drawings of everyday objects, animals, plants or food. First, two learning trials are presented. During each learning trial, the subject is instructed to look at and remember 24 line drawings. These learning trials are strictly intended for learning, no test score data is gathered during these trials. The line drawings of the learning trials are presented one by one with four seconds of exposure time to each line drawing, e.g. a bicycle (See trial 1 and 2 of Figure 1). Next, the immediate recognition trial is administered. During this trial, 24 visual associations are presented. In the visual associations the same 24 line drawings of the learning trials are presented actively interacting with another new line drawing e.g. the bicycle drawing has changed to a dog riding that same bicycle (See trial 3 of Figure 1). The subject is instructed to name both line drawings aloud, e.g. the bicycle as well as the dog, before pointing out the line drawing learned during the learning trial, e.g. the bicycle. By naming both line drawings aloud incidental learning of the visual associations is induced. Next, the delayed recognition trial is administered. During this trial, the same 24 visual associations are presented (See trial 4 of Figure 1) in exactly the same way as in the immediate recognition trial, but after a time interval of 15 minutes. During this time interval non-visual or non-memory tests were administered, which were part of the standard neuropsychological examination of the hospital. The immediate and delayed recognition trials are performance validity measures (Meyer, de Jonghe, Schmand, & Ponds, 2017). The immediate and delayed recognition trial scores range from 0 to 24. Next, the paired associate recall trial is administered. During this trial, the pictures from the learning trials are again presented one by one, e.g. the bicycle (See trial 5 of Figure 1), but now as cues for the subject to recall the associated line drawing, e.g. the dog. The paired associate recall trial scores range from 0 to 12; the total scores of this trial were

used in our study. Next, the free recall trial is administered. During this trial, the subject has to recall all line drawings that were previously shown, i.e. both line drawings in the visual associations, e.g. the bicycle and the dog. The free recall trial scores range from 0 to 48; the total scores of this trial were used in our study. Finally, the multiple-choice cued recognition trial is administered. During this trial, the last 12 line drawings from the learning trials are again shown one by one as cues, e.g. the bicycle, but now the associated line drawing, e.g. the dog, is also presented, which has to be recognized and pointed out among three distractors (See trial 7 of Figure 1). The multiple-choice cued recognition trial scores range from 0 to 12; the total scores of this trial were used in our study. During all trials, feedback is given with regard to correct or incorrect responses and an incorrect response is also corrected except during free recall.

Global cognitive impairment was measured with the Cognitive Screening Test (De Graaf & Deelman, 1991), which comprises a questionnaire about elementary factual knowledge and about orientation in time, place, and person. The scores range from 0 to 20, a cut-off score of less than 13 indicates cognitive impairment.

The Dutch version of the Rey Auditory Verbal Learning Test (Saan & Deelman, 1986) was used, in which the test administrator reads out loud a series of 15 words five times, which the subject is instructed to remember. After each presentation, the words have to be recalled. After a time interval of 20 minutes, the subject has to recall the same 15 words for the delayed free recall trial. The delayed free recall trial scores range from 0 to 15. Immediately after delayed free recall, the recognition trial is administered. During this trial, the test administrator reads out loud a series of 30 words that include the previously learned 15 words, and for each word the subject is instructed to state whether he recognizes it or not. The recognition trial scores range from 0 to 30.

Chapter 3

The level of education was scored with the Dutch scoring system Verhage (Verhage, 1964), which is similar to the International Standard Classification of Education 1997 scales (UNESCO, 2006). Scores range from one, i.e. unfinished elementary education to seven, i.e. university degree education.

| Trial | Sample of trial | Correct response |
|-------------------------------------|---|------------------|
| 1. Learning |  | - |
| 2. Learning |  | - |
| 3. Immediate recognition |  | Bicycle |
| 4. Delayed recognition |  | Bicycle |
| 5. Paired associate recall |  | Dog |
| 6. Free recall | - | Dog, Bicycle |
| 7. Multiple-choice cued recognition |   | Dog |

Figure 1. Sample of VAT-E trials.

Note. VAT-E = Visual Association Test-Extended.

Data analysis

In order to establish adequate study power we calculated the required number of subjects using G*Power for Window, version 3.1 (Faul, Erdfelder, Lang, & Bucher, 2007). Power ($1-\beta$) was set at 0.95, the level of significance at the α level of $p < .05$, one-tailed, and effect size was set at 0.9, as we expected to find large effect sizes (Cohen's d) (Meyer et al., 2017; Meyer et al., 2019). The Wilcoxon Signed-ranks test required a minimum of 16 test subjects and the t-test required a minimum of 15 test subjects.

We compared demographic characteristics between groups of healthy elderly controls, a-MCI patients, and mild AD patients, as appropriate. Also, we compared the Cognitive Screening Test total score and the Rey Auditory Verbal Learning Test delayed recall total score between these groups, as appropriate. To enable comparison within groups, we presented Rey Auditory Verbal Learning scores in percentage correct for each group. We corrected the Rey Auditory Verbal Learning score for chance responding, i.e. score * 0.50. To investigate if we could replicate previous research that indicated that verbal recognition memory improves after recall in AD (Bäckman et al., 2005), we compared the mean score of Rey Auditory Verbal Learning recognition corrected for chance responding with Rey Auditory Verbal Learning delayed free recall, as appropriate.

To analyze if the 24 VAT-E learning trial items were adequately learned, we calculated the mean, standard deviation, median, and range of scores for the VAT-E immediate recognition trial for all three groups. Furthermore, we compared the VAT-E subtest scores between healthy elderly controls, a-MCI patients, and AD patients, as appropriate.

To enable comparison within groups, we presented VAT-E scores in percentage correct scores for each group. This was done because the raw score of paired associate recall, free recall, and multiple-choice cued recognition each have a different range of scores, i.e. 0

to 24, 0 to 48, and 0 to 12 respectively. Also, for our within groups analyses, we corrected the VAT-E multiple-choice cued recognition scores for chance responding, i.e. score * 0.75. This was done because when four alternatives to choose from are presented, a subject will score 25 % correct if he is responding by chance alone. To correct for this, one has to multiply the obtained score by 0.75, therefore subtracting the chance score part of 25 % from the total score. We did not use this correction before because in our previous research (Meyer et al., 2019) we analyzed scores within each subtest.

To show the VAT-E score profile, we plotted the VAT-E subtest scores in percentage correct for each group in a line graph. The mean score of multiple-choice cued recognition corrected for chance responding was compared with the mean scores of VAT-E paired associate recall and VAT-E free recall within groups of healthy elderly controls, a-MCI patients or mild AD patients, as appropriate. In addition, these same comparisons were made separately for females and males in all groups, and for ‘young’ and ‘old’ subjects in all groups, i.e. as ‘young’ we classified all subjects younger than and including the median score of the group, and as ‘old’ we classified all subjects older than the median score of the group.

Results

Table 1 shows the demographic characteristics and test performances per group. Healthy elderly controls, a-MCI patients, and mild AD patients differed in gender ($\chi^2(2) = 14.39, p < .01$) and age ($F(2,101) = 15.87, p < .001$). Intermediate analyses showed that these differences did not affect VAT-E scores. The groups differed in Cognitive Screening Test score ($H(2) = 63.66, p < .001$) and Rey Auditory Verbal Learning Test delayed free recall score ($H(2) = 73.46, p < .001$). For a-MCI patients, the Rey Auditory Verbal Learning Test recognition score corrected for chance responding ($Mdn = 41.67$) was higher than the score for delayed free recall ($Mdn = 10.0$) ($z = -5.45, p < .001$). For mild AD patients, the Rey

Auditory Verbal Learning Test recognition score corrected for chance responding ($Mdn = 36.7$) was higher than the score for delayed free recall ($Mdn = 0$) ($z = -3.41, p < .001$).

The total scores of the VAT-E immediate recognition trial showed a strong ceiling effect for healthy elderly controls (100, 0, 100, 100-100), a-MCI patients (98.3, 2.92, 100, 91.7-100), and mild AD patients (97.4, 4.2, 100, 87.5-100). Healthy elderly controls, a-MCI patients, and AD patients differed in VAT-E scores for paired associate recall ($H(2) = 55.14, p < .001$), free recall ($H(2) = 71.18, p < .001$), and multiple-choice cued recognition ($H(2) = 51.82, p < .001$). For VAT-E paired associate recall, healthy elderly controls ($Mdn = 83.3$) scored higher than a-MCI patients ($Mdn = 20.8$) ($U = 174.0, p < .001$) and mild AD patients ($Mdn = 16.7$) ($U = 29.5, p < .001$); the scores of a-MCI patients and mild AD patients were not significantly different from one another ($U = 297.0, p = .111$). For VAT-E free recall, healthy elderly controls ($Mdn = 43.8$) scored higher than a-MCI patients ($Mdn = 10.4$) ($U = 31.0, p < .001$), and a-MCI patients scored higher than mild AD patients ($Mdn = 6.3$) ($U = 248.5, p < .05$). For VAT-E multiple-choice cued recognition, healthy elderly controls ($Mdn = 100.0$) scored higher than a-MCI patients ($Mdn = 83.30$) ($U = 295.0, p < .001$) and mild AD patients ($Mdn = 75.0$) ($U = 31.5, p < .001$); the scores of a-MCI patients and mild AD patients were not significantly different from one another ($U = 281.0, p = .064$).

The VAT-E score profile per group are shown in figure 2. For healthy elderly controls, the VAT-E score multiple-choice cued recognition corrected for chance responding ($Mdn = 75.0$) was higher than the VAT-E score for free recall ($Mdn = 43.8$) ($z = -5.56, p < .001, d = 2.7$). For a-MCI patients, the VAT-E score multiple-choice cued recognition corrected for chance responding ($Mdn = 62.48$) was higher than the VAT-E scores for paired associate recall ($Mdn = 20.8$) ($z = -4.91, p < .001, d = 1.4$) and for free recall ($Mdn = 10.4$) ($z = -5.65, p < .001, d = 3.5$), both with a large effect size. For mild AD patients, the VAT-E score multiple-choice cued recognition corrected for chance responding ($M = 49.3, SD = 18.9, Mdn$

= 56.3) was higher than the VAT-E scores for paired associate recall ($M = 18.6, SD = 18.9$) ($t(18) = 5.97, p < .001, d = 1.6$) and free recall ($Mdn = 6.3$) ($z = -3.72, p < .001, d = 2.7$), both with a large effect size.

Analyses showed that demographic differences in gender and age did not affect the VAT-E score profile of the groups. For healthy elderly controls, females and males scored higher on recognition than on free recall ($z = -5.05, p < .001; z = -2.37, p < .05$). For a-MCI patients, females and males scored higher on recognition than on paired associate recall ($z = -2.86, p < .01; z = -4.0, p < .001$) and on free recall ($t(18) = -10.49, p < .001; z = -4.20, p < .001$). For mild AD patients, females and males scored higher on recognition than on paired associate recall ($t(9) = -5.22, p < .01; t(8) = -3.59, p < .01$) and on free recall ($t(8) = -8.68, p < .001; t(8) = -4.87, p < .01$). For healthy elderly controls, ‘young’ and ‘old’ scored higher on recognition than on free recall ($z = -4.17, p < .001; z = -3.73, p < .001$). For a-MCI patients, ‘young’ and ‘old’ scored higher on recognition than on paired associate recall ($z = -3.80, p < .001; z = -3.20, p < .01$) and on free recall ($z = -4.02, p < .001; z = -4.02, p < .001$). For mild AD patients, ‘young’ and ‘old’ scored higher on recognition than on paired associate recall ($t(9) = -3.24, p < .05; t(8) = -6.12, p < .001$) and on free recall ($t(9) = -5.59, p < .001; t(7) = -6.49, p < .001$).

Table 1. Demographic characteristics and test performances per group.

| | Healthy elderly controls (<i>n</i> = 43) | a-MCI (<i>n</i> = 42) | AD (<i>n</i> = 19) |
|---|--|-----------------------------|------------------------------|
| Female (%) | 83.7 | 45.2 | 52.6 |
| Age | 72.7 (5.4) | 77.4 (5.2) | 80.7 (6.5) |
| Education | 4.6 (1.1) | 4.2 (1.1) | 4.3 (1.1) |
| VAT-E paired associate recall | 74.7 (20.0; 83.3; 20.8-100) | 29.0 (25.3; 20.8; 4.2-87.5) | 18.6 (18.9; 16.7; 0-75.0) |
| VAT-E free recall | 46.2 (14.7; 43.8; 18.8-83.3) | 12.2 (7.9; 10.4; 0-35.4) | 8.7 (9.5; 6.3; 0-33.3) |
| VAT-E multiple-choice cued recognition | 99.2 (2.4; 100; 91.7-100) | 77.4 (22.1; 83.3; 25.0-100) | 65.8 (25.2; 75.0; 8.3-100) |
| Cognitive Screening Test | 19.2 (1.2; 15-20) | 15.6 (2.2; 8-20) | 14.2 (2.7; 6.5-18) |
| Rey Auditory Verbal Learning Test delayed free recall | 56.1 (16.3; 53.3; 33.3-86.7) | 12.7 (13.0; 10.0; 0-40) | 4.6 (6.3; 0; 0-20) |
| Rey Auditory Verbal Learning Test recognition | 96.3 (4.1; 96.7; 86.7-100) | 79.3 (13.2; 83.3; 50-100) | 72.4 (13.7; 73.3; 46.7-96.7) |

Note. a-MCI = amnesic mild cognitive impairment; AD = Alzheimer's Disease; *n* = number of subjects; female (proportion of females); age in years (mean and standard deviation); education score based on the

Verhage system (mean and standard deviation); VAT-E = Visual Association Test-Extended total scores as percentage correct scores (mean, standard deviation, median, and range of scores); Cognitive Screening Test (mean, standard deviation, and range of scores); Rey Auditory Verbal Learning Test total scores as percentage correct scores (mean, standard deviation, median, and range of scores).

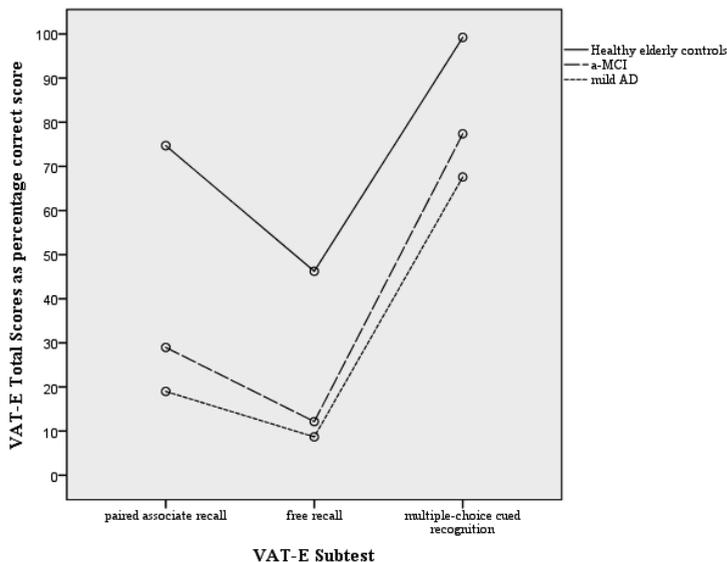


Figure 2. Score profiles of VAT-E total scores as percentage correct scores for healthy elderly controls ($n = 43$), a-MCI patients ($n = 42$), and AD patients ($n = 19$).

Note. a-MCI = amnesic mild cognitive impairment; AD = Alzheimer's Disease; n = number of subjects; VAT-E = Visual Association Test-Extended.

Discussion

This visual associative learning paradigm study based on incidental learning shows that episodic recognition memory is largely preserved when compared to recall in a-MCI or mild AD. While patients on average spontaneously recalled only 8 to 12 percent of VAT-E test items, they were actually able to remember, as evidenced by a 65 to 77 percent correct score during the recognition phase of the test. This score profile was similar across healthy elderly controls, a-MCI, and mild AD. These findings indicate that the episodic memory function in AD is not irreversibly lost in the early phase of the disease. This suggests a role for a retrieval impairment rather than solely a consolidation impairment.

This study adds to our previous findings on the VAT-E measures of memory (Meyer et al., 2019), in which we determined the degree of difficulty of these measures and found that

less difficult measures, i.e. measures that provide more retrieval support, are better able to retrieve memory as disease severity increases. In Meyer et al. (2019) we analyzed test performances for each measure separately in relation to global cognitive impairment to establish the amount of floor effect, with larger floor effects indicating increasingly difficult measures. However, that psychometrical analysis left unanswered what this means with regard to the supposed consolidation impairment as stated in the core diagnostic criteria for probable AD (Dubois et al., 2007). To answer this an analysis of the score profiles of the groups was necessary. Therefore, contrary to Meyer et al. (2019), in our present study we examined test score profiles within groups. Also, contrary to Meyer et al. (2019) in our present study we included a-MCI and mild AD patients not only based on clinical symptoms, but also based on medial temporal lobe atrophy supportive of probable AD as evidenced by structural magnetic resonance imaging (Scheltens et al., 1992), thereby investigating if a-MCI or mild AD patients are able to recognize learned material despite hippocampal atrophy.

A strong point of our study is the use of multiple memory measures, i.e. paired associate recall, free recall, and multiple-choice cued recognition, of a well-validated episodic memory test, i.e. the VAT-E. A second strong point is our cross-sectional design in which we included healthy elderly controls, a-MCI patients, and mild AD patients, representing the continuum of the early stages of Alzheimer's disease. A third strong point is the use of medial temporal lobe atrophy supportive for the diagnosis of probable AD as evidenced by magnetic resonance imaging (Scheltens et al., 1992), thus findings of preserved recognition memory in mild AD are not only based on syndromal diagnostic criteria, but also on demonstrated cerebral atrophy. Since we used this surrogate biomarker of AD, we consider it highly probable that our a-MCI group represents the prodromal stage of AD.

Our findings may have implications for clinical assessment of a-MCI or AD patient. Typically, cognitive deficits in these patients are measured with tests of free recall. Also,

impaired free recall is often contrasted with recognition during neuropsychological assessment of a patient. Since our findings show that preserved recognition compared to recall may be compatible with AD being the correct diagnosis, this may have implications for accurate assessment of the memory dysfunction in AD is. Rather than a tool to differentiate AD from other dementias, the extent of preserved recognition memory may be an indicator of the severity of the episodic memory impairment.

Health care professionals could employ memory tools to improve patients' daily memory functioning based on visual associations and retrieval support. Previous research shows that patients with a-MCI and early stage AD can learn to use cues that are based on associate learning and recognition (Cavallo et al., 2013; Hampstead et al., 2019). For example, if a patient still benefits from cues based on visual associative learning, health care professionals could help patients to learn to associate daily living activities with icons, which could be used to light up on a planning board according to a timetable that was agreed upon by the patient. Also, if a patient still benefits from retrieval support based on recognition, health care professionals could use multiple-choice questions instead of free recall questions when talking to the patient.

Our conclusion that AD is characterized by a retrieval impairment rather than by a consolidation impairment seems to contradict previous findings. However, although previous studies reported that decreased sensitivity to semantic cueing in AD can differentiate between subjects, i.e. distinguish AD patients from healthy elderly controls (Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994; Tounsi et al., 1999) or from non-AD disorders such as progressive supranuclear palsy (Pillon et al., 1994), these studies also indicated that within the AD groups episodic memory based on semantic cueing improved after free recall. Furthermore, these previous studies only studied the effect of cueing and not recognition. However, there are many studies that show that verbal recognition memory in patients with a-

MCI or mild AD is more intact than free recall (Bäckman et al., 2005). A finding that is replicated in our study, as Rey Auditory Verbal Learning Test scores improved on recognition after delayed free recall. In addition, we show that patients with a-MCI or mild AD can indeed recognize previously learned material despite a deficit on recall and that not only verbal, but also visual recognition memory is partially intact. Specifically, our study is the first to show that when patients with a-MCI or mild AD incidentally learn a picture that is linked with another picture by means of an interaction, their episodic recognition memory improves after recall.

There has been disagreement in the field regarding the function of the medial temporal lobe (Montaldi & Mayes, 2010). One view is that these structures, i.e. the hippocampal region, and the entorhinal, perirhinal, and parahippocampal cortices (Squire, Stark, & Clark, 2004), contribute to episodic memory in a similar fashion (Squire & Bayley, 2007). However, our patients all have medial temporal lobe atrophy supportive for the diagnosis of probable AD (Scheltens et al., 1992), and nevertheless their memory improves after recall when tested with recognition. Since medial temporal lobe atrophy supportive for the diagnosis of probable AD is defined as volume loss of the hippocampal region and the entorhinal cortex (Dubois et al., 2007), this suggest a functional differentiation for the other structures of the medial temporal lobe. A possible explanation is postulated by Montaldi and Mayes (2010), namely that the perirhinal and parahippocampal cortices are critical for visual episodic recognition memory, whereas the hippocampal region is critical for recall, and that the perirhinal cortex is the mediating structure to subsequently increase activity in associated cortices such as frontal, cingulate, and parietal cortices, and the thalamus (Montaldi et al., 2006).

Our findings add to the theoretical framework of knowledge about the memory functioning of patients with AD. We show that, in addition to prior literature on this topic (Bäckman et al., 2005), a visual associative learning paradigm based on incidental learning is

effective in establishing largely preserved episodic memory in AD. The visual associations enhance the engagement with the test material by means of dual coding, the interaction, and the unexpectedness of this interaction. Firstly, in contrast to words, pictures are more often encoded visually as well as verbally, i.e. dual coding (Paivio, 1991). Secondly, the interaction depicted in the visual association strongly links the two items presented in the picture (Lindeboom et al., 2002). Thirdly, the unexpectedness of this interaction stimulates elaboration (Birngruber, Schröter, & Ulrich, 2014; Ulrich, Nitschke, & Rammsayer, 2006). In addition, by using incidental learning the clinician has more control over the way the patient processes the material, because the patient cannot use a mnemonic strategy that is unknown to the clinician.

A limitation of our study is that we did not use a longitudinal design. However, although we did not have follow-up data on memory test performances, we consider our findings to be robust, due to the rigorous diagnostic process within the cross-sectional design. Nevertheless, future research should use follow-up data, preferably adding results of patients with more advanced disease stages, in which we would expect a gradual loss of the effect on recognition as disease stages become more severe.

The purpose of our study was to improve the assessment of memory dysfunction in AD. Therefore, we investigated preserved episodic recognition memory using incidental learning of visual associations in patients with a-MCI or mild AD. We show that the episodic memory impairment in AD is not absolute and may be characterized as a retrieval impairment rather than a consolidation impairment, indicating that preserved recognition compared to recall may be compatible with AD being the correct diagnosis. In addition, as far as we know, the VAT-E is the first test that uses incidental learning of visual associations. Since the VAT-E combines recall and recognition trials in one single test, the clinician can see at a glance if

patients improve on recognition after they showed a deficit on recall. Thus, the VAT-E can potentially be a valuable tool for clinical assessment.

Acknowledgments

We thank the department of geriatric medicine of the Northwest Medical Center location Alkmaar, the Netherlands, for enabling us to do our research.

Disclosure Statement

S.R.A.M and J.F.M.J are the authors of the Visual Association Test-Extended.

Funding

This work was supported by Foreest Northwest Medical School [grant number FIO1308].

References

- Ally, B. A. (2012). Using pictures and words to understand recognition memory deterioration in amnesic mild cognitive impairment and Alzheimer's disease: A review. *Current Neurology and Neuroscience Reports*, *12*, 687-694.
- Bäckman, L., Jones, S., Berger, A., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, *19*, 520-531.
- Birngruber, T., Schröter, H., & Ulrich, R. (2014). Duration perception of visual and auditory oddball stimuli: Does judgement task moderate the temporal oddball effect? *Attention, Perception, & Psychophysics*, *76*, 814-828.
- Cavallo, M., Cavanna, A. E., Harciarek, M., Johnston, H., Ostacoli, L., & Angilletta, C. (2013). "Keep up the good work!": a case study of the effects of a specific cognitive

- training in Alzheimer's disease. *Neurocase*, 19, 542-552.
- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- De Graaf, A. & Deelman, B.G. (1991). *Cognitive Screening Test. Manual for administration and scoring*. Lisse: Swets en Zeitlinger bv.
- Deweert, B., Lehericy, S., Pillon, B., Baulac, M., Ciras, J., Marsault, C., ... Dubois, B. (1995). Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. *Journal of Neurology, Neurosurgery, and Psychiatry*, 58, 590-597.
- Dierckx, E., Engelborghs, S., Raedt, R. de, Buggenhout, M. van, Deyn, P. P. de, Verté, D., & Ponjaert-Kristofferen, I. (2009). Verbal cued recall as a predictor of conversion to Alzheimer's disease in mild cognitive impairment. *International Journal of Geriatric Psychiatry*, 24, 1094-1100.
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ...Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, 6, 734-746.
- Faul, F., Erdfelder, E., Lang, A.-G., & Bucher, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fuchs, A., Wiese, B., Altiner, A., Wollny, A., & Penzek, M. (2011). Cued recall and other cognitive tasks to facilitate dementia recognition in primary care. *Journal of American Geriatric Society*, 60, 130-135.
- Gifford, K. A., Phillips, J. S., Samuels, L. R., Lane, E. M., Bell, S. P., Liu, D., ... for the Alzheimer's Disease Neuroimaging Initiative (2015). Associations between verbal

learning slope and neuroimaging markers across the cognitive aging spectrum. *Journal of the International Neuropsychological Society*, 21, 455-467.

Hampstead, B. M., Stringer, A. Y., Stilla, R. F., & Sathian, K. (2019). Mnemonic strategy training increases neocortical activation in healthy older adults with mild cognitive impairment. *International Journal of Psychophysiology*. Advanced online publication. doi.org/10.1016/j.ijpsycho.2019.04.011.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, 36, 1351-1359.

Kessels, R. P. C., Rijken, S., Joosten-Weyn Banningh, L. W. A., van Schuylenborgh-van Es, N., & Olde Rikkert, M. G. M. (2009). Categorical spatial memory in patients with mild cognitive impairment and alzheimer dementia: Positional versus object-location recall. *Journal of the International Neuropsychological Society*, 16, 200-204.

Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.

Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73, 126-133.

Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended. Handleiding. [Visual Association Test-Extended. Manual]*. Amsterdam: Hogrefe Publishing bv.

Meyer, S. R. A., de Jonghe, J. F. M., Schmand, B., & Ponds, R. W. H. M. (2017). The Visual Association Test-Extended: A cross-sectional study of the performance validity measures. *The Clinical Neuropsychologist*, 31, 789-813.

Meyer, S. R. A., de Jonghe, J. F. M., Schmand, B., & Ponds, R. W. H. M. (2019). Visual

- associations to retrieve episodic memory across healthy elderly, mild cognitive impairment, and patients with Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, 26, 447-462.
- Meyer, S. R. A., Spaan, P. E. J., Boelaarts, L., Ponds, R. W. H. M., Schmand, B., & de Jonghe, J. F. M. (2016). Visual associations cued recall A paradigm for measuring episodic memory decline in Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, 23, 566-577.
- Montaldi, D. & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*, 20, 1291-1314.
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, 16, 504-520.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255-287.
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11, 137-138.
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183-194.
- Petersen, R. C., Doody, R., Kurz, A. Mohs, R. C., Morris, J.C., Rabins, P. V., ... Winbald, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985-1992.
- Petersen, R. C., Smith, G. E., Ivnik, R. J., Kokmen, E., & Tangalos, E. G. (1994). Memory function in very early Alzheimer's disease. *Neurology*, 44, 867-872.
- Pillon, B., Deweer B., Michon, A., Malapani, C., Agid, Y., & Dubois, B. (1994). Are explicit memory disorders of progressive supranuclear palsy related to damage to striatofrontal circuits? Comparison with Alzheimer's, Parkinson's, and Huntington's disease.

Neurology, 44, 1264-1270.

Reesink, F. E., Lemstra, A. W., van Dijk, K. D. Berendse, H. W., van de Berg, W. D. J. Klein, M., ... van der Flier, W. M. (2010). CSF α -synuclein does not discriminate dementia with Lewy bodies from Alzheimer's disease. *Journal of Alzheimer's Disease*, 22, 87-95.

Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature*, 531, 508-512.

Saan, R. J. & Deelman, B. G. (1986). *15-Woorden Test A en B* [15-Word Test A and B]. In Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (Eds.) (2012). *Handboek neuropsychologische diagnostiek* [Handbook neuropsychological assessment], 2nd ed. (p.p. 267-282). Amsterdam: Pearson Assessment and Information B.V.

Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H.C., Vermersch, P., Kuiper, M., ... Kuiper, M. (1992). Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery, and Psychiatry*, 55, 967-972.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6, 156-163.

Skinner, E. I. & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologica*, 45, 2163-2179.

Squire, L. R. & Bayley P. J. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology*, 17, 185-196.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279-306.

Tounsi, H., Deweer, B., Ergis, A., Linden, M. van der, Pillon, B., Michon, A., et al. (1999).

- Sensitivity to semantic cueing: An index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 13, 38-46.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21, 437-484.
- Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research*, 70, 77-87.
- UNESCO (2006). *International Standard Classification of Education: ISCED 1997 (re-edition)*. Montreal: UNESCO-UIS. Retrieved May 31, 2018, from http://www.uis.unesco.org/education/information/nfsunesco/doc/isced_1997.htm.
- Van der Vlies, A. E., Verwer, N. A., Bouwman, F. H., Blankenstein, M. A., Klein, M., Scheltens, P., & van der Flier, W. M. (2009). CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology*, 72, 1056-1061.
- Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar* [Intelligence and age: Research within Dutch participants aged from 12 to 77 years old]. Dissertation. Assen: Van Gorcum.
- World Health Organization (2017). *Dementia fact sheet updated December 2017*. Retrieved January 11, 2018, from <http://www.who.int/mediacentre/factsheet/fs362/en/>.

4

Chapter 4

VISUAL ASSOCIATIONS TO RETRIEVE EPISODIC MEMORY ACROSS HEALTHY ELDERLY, MILD COGNITIVE IMPAIRMENT, AND PATIENTS WITH ALZHEIMER'S DISEASE

Sascha R.A. Meyer, Jos F.M. de Jonghe, Ben Schmand, and Rudolf W.H.M. Ponds

Aging, Neuropsychology, and Cognition, 2019; 26: 447-462

Abstract

Episodic memory tests need to determine the degree to which patients with moderate to severe memory deficits can still benefit from retrieval support. Especially in the case of Alzheimer's disease (AD), this may support health care to be more closely aligned with patients' memory capacities. We investigated whether the different measures of episodic memory of the Visual Association Test-Extended (VAT-E) can provide a more detailed and informative assessment on memory disturbances across a broad range of cognitive decline, from normal to severe impairment as seen in AD, by examining differences in floor effects. The VAT-E consists of 24 pairs of black-and-white line drawings. In a within-group design, we compared score distributions of VAT-E subtests in healthy elderly controls, mild cognitive impairment (MCI), and AD ($n = 144$), as well as in relation to global cognitive impairment. Paired associate recall showed a floor effect in 41% of MCI patients and 62% of AD patients. Free recall showed a floor effect in 73% of MCI patients and 84% of AD patients. Multiple-choice cued recognition did not show a floor effect in either of the patient groups. We conclude that the VAT-E covers a broad range of episodic memory decline in patients. As expected, paired associate recall was of intermediate difficulty, free recall was most difficult, and multiple-choice cued recognition was least difficult for patients. These varying levels of difficulty enable a more accurate determination of the level of retrieval support that can still benefit patients across a broad range of cognitive decline.

Introduction

Cognitive impairment is measured with episodic memory tests, but it is unclear whether such tests are clinically meaningful in other ways, particularly in determining if patients with Alzheimer's disease (AD) can still benefit from retrieval support. Ideally, this should be determined across a broad range of cognitive decline, so that health care can be more closely aligned with a patient's memory capacities by providing both health care professionals and informal caregivers with memory tools that improve the patient's daily memory functioning.

In clinical practice, most commonly used episodic memory tests are based on free recall paradigms, which are primarily used to distinguish between normal and abnormal, dementia-related memory functioning, ignoring more subtle differences in the level of intact memory processing across the broad range of global cognitive impairment in dementia patients. A more informative approach would be to use tasks with varying levels of retrieval support to assess the strength of memory processing in a more detailed way. Earlier research has shown that test paradigms that provide retrieval support by means of cued recall might improve retrieval of test material in patients with mild cognitive impairment (MCI) or AD (e.g. Greenaway et al., 2006; Tounsi et al., 1999). The retrieval of test material in these patients might further be improved if even more retrieval support is provided by using recognition paradigms (Bäckman, Jones, Berger, Laukka, & Small, 2005; Lowndes et al., 2008). In addition, research showed that free recall can be too difficult for patients with MCI or AD, thus resulting in floor effects on tests, whereas cued recall subsequently improved retrieval of episodic memory in these patients (Diesfeldt, Prins, & Lauret, 2017; Meyer et al., 2016). Therefore, several test paradigms with varying difficulty levels should be employed to cover a broad range of memory decline in these patients, and in order to enable comparison between test paradigms, these paradigms should be combined in one single test.

A well-known example of such a test is the California Verbal Learning Test (Delis, Kaplan, Kramer, & Ober, 1987), which combines free recall, cued recall and recognition paradigms. Research has shown that these measures can differentiate within a group of AD patients based on severity of cognitive impairment (Fox, Olin, Erblich, Ippen, & Schneider, 1998), thus covering a broad range of memory decline.

Since the California Verbal Learning Test is based on verbal stimuli, a similar test based on visual stimuli might be clinically useful to enable language-independent assessment of patient groups, such as dyslexic or non-native speakers. In addition, tests based on visual stimuli are particularly suitable for covering a broad range of memory decline because pictures are easier to remember than words (Shepard, 1967) and because this so-called picture superiority effect is applicable to normal controls for free recall (Paivio, Rogers, & Smythe, 1968), paired associate recognition (Hockley, 2008), and recognition (Shepard, 1967), as well as to patients with MCI or AD for recognition (Alley, Gold, & Budson, 2009). However, current visual memory tests often use abstract designs as learning material, e.g. Wechsler Memory Scale-IV subtest Visual Reproduction (PsychCorp, 2009; Wechsler, 1945), Rey's Complex Figure Test (Osterreith, 1944; Rey, 1941), and Benton Visual Retention Test (Sivan, 1992). Because these abstract designs have to be recalled by drawing the design from memory, this may complicate the interpretation since failure may be due to visuoconstructive disability. Also, since functioning in daily life mostly does not concern abstract designs, such tests seem less applicable to patients' daily memory functioning. Moreover, because the memory deficit in AD has mostly been studied by means of word list learning (Lezak, Howieson, Bigler, & Tranel, 2012), a test that is based on pictures related to daily life may be a valuable addition for studying memory in MCI and AD.

The Visual Association Test (VAT; Lindeboom & Schmand, 2003; Lindeboom, Schmand, Meyer, & de Jonghe, 2014) is an episodic memory test based on pictures of

everyday objects, animals, plants, or food. Various studies showed that the VAT has a high validity in predicting progression to AD (Dierckx et al., 2009; Fuchs, Wiese, Altiner, Wollny, & Pentzek, 2011; Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002; Reesink et al., 2010; Van der Vlies et al., 2009). We previously extended the VAT by doubling the number of test items and adding a free recall condition (Meyer, de Jonghe, Schmand, & Ponds, 2017), thus constructing the measures of memory of the Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017), i.e. paired associate recall, free recall and multiple-choice cued recognition, which, in conjunction with each other, may be suitable for covering a broad range of memory decline.

The VAT-E is based on visual associations shown during an incidental learning task (Craik & Lockhart, 1972; Tulving & Madigan, 1970). By means of the visual association, a picture that the patient is asked to remember is encoded together with a specific cue, which subsequently can be presented to maximize retrieval. By varying in the level of retrieval support, also the level of difficulty of the measures of memory varies, which in turn enables a broad range of memory decline to be covered. The VAT-E consists of pairs of pictures that are associated through an interaction, e.g. a dog is depicted riding a bicycle (See Figure 1). The first measure of memory, paired associate recall, is thought to be of intermediate difficulty, as it presents a cue, e.g. a picture of a bicycle, to facilitate recall (See Figure 1). The second measure of memory, free recall, is thought to be the most difficult, as no retrieval support is presented. The third measure of memory, multiple-choice cued recognition, is thought to be the least difficult, as it not only presents a cue, e.g. a picture of a bicycle, but also facilitates recognition by presenting the target and three alternatives (See Figure 1).

The purpose of our study was to investigate if the VAT-E covers a broad range of memory decline by examining the difficulty levels of the measures of episodic memory in patients with MCI or AD across the range of severity of global cognitive impairment. These

varying levels of difficulty would enable a more accurate determination of the level of retrieval support that can still benefit patients across a broad range of cognitive decline. The difficulty level of a memory test can be evaluated by examining floor effects, with larger floor effects indicating increasingly difficult measures (Cohen, Swerdlik, & Sturman, 2013). We expected the more difficult measures to show larger floor effects in patient groups with more severe global cognitive impairment. We hypothesized that the VAT-E measures of memory cover a broad range of cognitive decline in patients with MCI or AD, since they do not only establish, by means of free recall, whether a patient has an episodic memory impairment, but also give information, by means of paired associate recall and multiple-choice cued recognition, about the degree to which a patient can still benefit from retrieval support. We expected that paired associate recall would show an intermediate floor effect, as it is of intermediate difficulty for patient groups, while free recall was expected to show a large floor effect, as it is the most difficult task for patient groups, and multiple-choice cued recognition was expected to show no floor effect, as it is the least difficult task for patient groups.

Methods

Participants

Our study sample consisted of 44 healthy elderly controls, 74 MCI patients, and 26 mild AD patients. The MCI group consisted of 59 patients with amnesic MCI and 15 with amnesic MCI multiple domain (Petersen, 2004). Hearing and vision was intact or corrected to normal for all participants. We recruited patients from the outpatient memory clinic at the department of geriatric medicine of the Northwest Medical Center, a large teaching hospital in Alkmaar, the Netherlands. Healthy elderly controls were recruited amongst the informal caregivers of the participating outpatients, all of whom were close family members or spouses of the patients. Since all informal caregivers cared for the patients on a regular basis, they were

extensively interviewed by the clinical neuropsychologist regarding the medical history of the patient. If the informal caregivers appeared clinically normal, had no complaints of memory, and were generally unaffected in their activities of daily living, they were asked to participate in the study as healthy elderly controls.

Participants were included if they were 65 years or older. Participants were excluded if they had severe traumatic brain damage or brain disease (such as brain tumor, epilepsy, multiple sclerosis, Parkinson's disease), a severe psychiatric disorder, delirium or a history of alcohol or drug abuse. Healthy elderly controls were included if they scored above published cut-off scores, i.e. 13 or higher, on the Cognitive Screening Test (CST; De Graaf & Deelman, 1991). Healthy elderly controls underwent a screening interview concerning the exclusion criteria and were excluded, if appropriate. In addition, they were excluded if their Rey Auditory Verbal Learning Test delayed recall score was at percentile 5 or below corrected for gender, age, and education.

MCI and AD diagnoses were based on internationally accepted criteria (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision; American Psychiatric Association, 2000; Dubois et al., 2007; Petersen et al., 2001). The patients' diagnostic program included a comprehensive geriatric assessment, laboratory results, magnetic resonance imaging or computed tomography scanning, and a standardized neuropsychological examination. The final MCI or AD diagnosis was made after discussion of all clinical data by a multidisciplinary team consisting of five geriatricians, a neurologist, a clinical neuropsychologist and a geriatric psychiatrist. The study was approved by the local medical ethical committee. Written informed consent was obtained from all participants.

Design and procedure

In a within-group design, we compared VAT-E scores for paired associate recall, free recall, and multiple-choice cued recognition in groups of healthy elderly controls, MCI patients, and AD patients, in relation to the participants' severity of global cognitive impairment as measured with the CST. As an additional baseline measure, all participants took the Rey Auditory Verbal Learning Test (Rey, 1964; Schmidt, 1996) delayed recall trial. For patients, administration of the VAT-E, the CST, and the Rey Auditory Verbal Learning Test was part of the standardized neuropsychological examination of the hospital, which besides these tests, consisted of several tests and questionnaires, i.e. Cambridge Cognitive Examination subtest perception (Roth, Huppert, Mountjoy, & Tym, 1998), Animal Gnosis-card, Amsterdam Dementia-Screening Test subtests orientation, copy, and meander (Lindeboom & Jonker, 1989), Trail Making Test (Reitan & Wolfson, 1985), Stroop Colour-Word Test (Hammes, 1971), Clock drawing, Groninger Intelligence Test 2 subtests fluency animals, fluency professions, word meaning, and additional subtests if appropriate (Luteijn & Barelds, 2004), Rey's Complex Figure Test subtests copy and delayed recall, Geriatric Depression Scale (Sheik & Yesavage, 1986), Informant Questionnaire on Cognitive Decline in the Elderly (Jorm & Jacomb, 1989), Neuropsychiatric Inventory (Cummings et al., 1994), Lawton Instrumental Activities of Daily Living (Lawton & Brody, 1969), KATZ activities of daily living (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963), Clinical Dementia Rating (Hughes, Berg, Danziger, Coben, & Martin, 1982; Morris, 1993), and subtests of the Behavioural Assessment of the Dysexecutive Syndrome if appropriate (Wilson, Alderman, Burgess, Emslie, & Evans, 1996). Since healthy elderly controls were the informal caregivers of the patients, they did not undergo the standardized neuropsychological examination but only the VAT-E, the CST, and the Rey Auditory Verbal Learning Test. To avoid interference between these tests, they were administered in a fixed order, i.e. CST, VAT-E, and Rey Auditory

Verbal Learning Test, and during the time interval between the immediate and delayed conditions of the VAT-E and the Rey Auditory Verbal Learning Test, participants were requested to complete questionnaires. However, if, in spite of this, there would have been interference, this would be equal for all participants due to the fixed order. Healthy elderly controls completed these tests at the same time as the patients, but in a separate room.

Since the VAT-E was not used to determine the patients' diagnosis, we removed the VAT-E score sheet from the patient's hospital dossier to ensure that the clinical neuropsychologist was blinded to the VAT-E score. The CST could not be removed from the patient's hospital dossier, as it is part of the hospital's standardized neuropsychological examination.

Measures

The Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017) is based on items of the Visual Association Test (VAT; Lindeboom & Schmand, 2003; Lindeboom, Schmand, Meyer, & de Jonghe, 2014) and consists of the joined VAT 12-item versions AB and CD. Only the CD version has a four-alternative forced-choice cued recognition trial. This version was used for the multiple-choice cued recognition trial of the VAT-E.

The VAT-E measures of episodic memory, paired associate recall, free recall, and multiple-choice cued recognition are based on 24 pictures of black-and-white line drawings. The stimuli are learned in two learning trials and two recognition trials, i.e. immediate recognition and delayed recognition. First, each learning trial presents the same 24 pictures one by one, e.g. a bicycle, which the subject is instructed to look at and remember. In the immediate recognition trial, each picture from the learning trials is linked with an associated picture through an interaction, e.g. a dog riding a bicycle (See Figure 1). The subject is instructed to name both pictures, thus inducing incidental learning of the association, and to

point out the picture from the learning trials, e.g. the bicycle. After 15 minutes, delayed recognition is administered. During this time interval, subjects can complete non-visual tests, non-memory tests, or questionnaires. Delayed recognition is administered in the same way as immediate recognition, i.e. the subject is instructed to name both pictures and point out the picture from the learning trials. In this way, the subject is being tested twice on recall of the learning trial pictures. The immediate and delayed recognition trials are employed as performance validity measures (Meyer, de Jonghe, Schmand, & Ponds, 2017).

Next, the paired associate recall trial is administered, in which the pictures from the learning trials are shown again, but now one by one as cues, e.g. the bicycle (See Figure 1), and the subject is instructed to recall the associated picture, e.g. the dog. The paired associate recall scores range from 0 to 24. Next, the free recall trial is administered, in which the subject is instructed to recall all cues and all associated pictures without being shown any of the pictures. The free recall scores range from 0 to 48. Finally, the multiple-choice cued recognition trial is administered, in which the last 12 pictures from the learning trials are shown again one by one as cues, and the subject is instructed to recognize and point out the associated picture out of four alternatives (See Figure 1). The multiple-choice cued recognition scores range from 0 to 12.

During all trials, the test administrator gives feedback regarding correct or incorrect responses on each item. If, during paired associate recall, the subject gives an incorrect answer, the test administrator gives the correct answer so that the exposure to the cue and the associated picture is equal for all subjects before the free recall trial. An incorrect response is also corrected during multiple-choice cued recognition, but not during free recall. The VAT-E measures of memory are highly predictive of MCI and AD, with areas under the curve from .81 to .99 for MCI, and from .97 to .99 for AD (Meyer & de Jonghe, 2017). Normative data for the VAT-E measures of memory were established (Meyer & de Jonghe, 2017).

The severity of global cognitive impairment was measured with the Cognitive Screening Test (CST; De Graaf & Deelman, 1991), which consists of 20 questions relating to orientation in time, place, and person, and elementary factual knowledge, e.g. “when was the second World War?”. The scores range from 0 to 20. High scores indicate lack of impairment and lower scores indicate more severe impairment. The CST is a Dutch version of the short portable mental status questionnaire (Pfeiffer, 1975) and the mental status questionnaire (Kahn, Goldfarb, Pollack & Peck, 1960). The CST correlates 0.86 with the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975; Dautzenberg, Schmand, Vriens, Deelman & Hooijer, 1991). In a study based on a large sample, $n = 4051$, the CST was compared to the Mini-Mental State Examination and was found to be equally predictive of dementia, with areas under the curve of .91 (95% CI = .88-.93) and .92 (95% CI = .90-.94) respectively (Schmand, Deelman, Hooijer, Jonker, & Lindeboom, 1996). In another study based on a large sample, $n = 761$, the CST was also highly predictive of dementia, with an area under the curve of .94 (Van Toutert, Diesfeldt, & Hoek, 2016). The reliability of the CST is high, with Cronbach’s alpha being 0.89 (Dautzenberg et al., 1991) and test-retest reliability being 0.80 (Schmand et al., 1996).

This study used the Dutch version of the Rey Auditory Verbal Learning Test (Saan & Deelman, 1986), which presents a list of 15 unrelated words in five consecutive learning trials. After each presentation and after a 20-minute time interval, i.e. delayed recall, these words have to be recalled. The delayed recall scores range from 0 to 15.

The level of education was scored with a Dutch system (Verhage, 1964) that is similar to the International Standard Classification of Education 1997 scales (UNESCO, 2006). The scores range from one to seven, i.e. pre-primary education, primary education, lower secondary education, upper secondary education, post-secondary non-tertiary education, first stage tertiary education, and second stage tertiary education respectively.



Figure 1. Sample of visual association (left), cue (middle), and multiple-choice answer (right).

Data Analysis

The effects of demographic characteristics on VAT-E paired associate recall, free recall, and multiple-choice cued recognition were analyzed by means of Pearson's chi-square test, Kruskal-Wallis tests, Mann-Whitney tests, t-tests and Spearman's and Pearson's correlations, as appropriate. Also, we analyzed differences between groups on VAT-E paired associate recall, free recall, and multiple-choice cued recognition, as appropriate. VAT-E scores of healthy elderly controls, MCI patients, and AD patients were pooled. VAT-E scores were plotted against the range of global cognitive impairment, i.e. the participants' CST scores. Hierarchical multiple regression analysis was used to determine whether a linear or quadratic model fitted the data best. To predict VAT-E test performances, we entered the CST score, representing a linear model, at step 1, and entered the squared CST score, representing a quadratic model, at step 2. This resulted either in a linear model as best fit for the data if the total variance explained was significantly increased by the CST score alone or a quadratic model as best fit for the data if the total variance explained was significantly increased by the CST score together with the squared CST score.

If the best fit for the data is a quadratic model, this means that VAT-E scores level off as severity of global cognitive impairment increases, thus indicating a floor effect. The range of such a floor effect was defined by scores below the VAT-E score that was obtained from entering the minimum CST score into the quadratic equation and calculating the outcome.

Subsequently, we calculated the percentage of subjects scoring within the range of the floor effect. Preliminary analyses showed that the assumption of multicollinearity was violated, but this was as expected, because the two predictors in the model are based on the same variable, i.e. the CST score. Also, the assumption of linearity was violated, as expected, for the quadratic model.

Results

The demographic characteristics and test performances per group are shown in table 1. The groups differed in gender ($\chi^2(2) = 15.99, p < .001$) and age ($H(2) = 18.74, p < .001$). These differences had no effect on VAT-E scores. For paired associate recall, healthy elderly controls ($Mdn = 20.0$) scored higher than MCI patients ($Mdn = 6.0$) ($U = 398.5, p < .001$), and MCI patients scored higher than AD patients ($Mdn = 3.5$) ($U = 636.0, p < .05$). For free recall, healthy elderly controls ($Mdn = 21.0$) scored higher than MCI patients ($Mdn = 5.0$) ($U = 52.0, p < .001$), and MCI patients scored higher than AD patients ($Mdn = 3.0$) ($U = 539.5, p < .01$). For multiple-choice cued recognition, healthy elderly controls ($Mdn = 12.0$) scored higher than MCI patients ($Mdn = 10.5$) ($U = 668.0, p < .001$), and MCI patients scored higher than AD patients ($Mdn = 7.5$) ($U = 543.0, p < .01$).

Table 1. Demographic characteristics and test performances per group.

| | Healthy elderly controls (<i>n</i> = 44) | MCI (<i>n</i> = 74) | AD (<i>n</i> = 26) |
|--|--|-------------------------|-------------------------|
| Female (%) | 84.1 | 47.3 | 53.8 |
| Age | 73.0 (5.6) | 76.5 (5.6) | 79.5 (6.7) |
| Education | 4.6 (1.1) | 4.2 (1.1) | 4.3 (1.0) |
| VAT-E paired associate recall | 18.1 (4.8; 23.3; 5-24) | 7.9 (6.5; 42.6; 0-23) | 4.3 (4.3; 18.7; 0-18) |
| VAT-E free recall | 22.8 (8.0; 64.1; 9-48) | 6.0 (4.0; 15.6; 0-17) | 3.6 (4.2; 17.8; 0-16) |
| VAT-E multiple-choice cued recognition | 11.9 (0.3; 0.1; 11-12) | 9.6 (2.7; 7.3; 2-12) | 7.2 (3.5; 11.9; 1-12) |
| CST | 19.2 (1.2; 1.3; 15-20) | 16.0 (2.0; 4.2; 8-20) | 13.6 (2.7; 7.5; 6.5-18) |
| Rey Auditory Verbal Learning Test delayed recall | 8.5 (2.5; 6.3; 5-13) | 2.0 (1.8; 3.3; 0-6) | 0.9 (1.5; 2.3; 0-6) |

Note. MCI = mild cognitive impairment; AD = Alzheimer's Disease; *n* = number of subjects; female (proportion of females); age in years (mean and standard deviation); education score based on the Verhage system (mean and standard deviation); VAT-E = Visual Association Test-Extended (mean, standard deviation, variance, and range of scores); CST = Cognitive Screening Test (mean, standard deviation, variance and range of scores); Rey Auditory Verbal Learning Test (mean, standard deviation, variance and range of scores).

Table 2 and Figures 2-4 show the test performances of VAT-E paired associate recall, free recall, and multiple-choice cued recognition in relation to severity of global cognitive impairment. For paired associate recall, the explained variance in test performance was increased by the squared CST scores. Thus, the best fit for the paired associate recall data was a quadratic regression model ($F(2, 141) = 46.87, p < .001$) (See Figure 2), which indicated a floor effect. The range of this floor effect comprised scores of 4 or below, which were scored by 41% of MCI patients and 62% of AD patients. No healthy elderly controls scored within the range of this floor effect. For free recall, the explained variance in test performance was increased by the squared CST scores. Thus, the best fit for the free recall data was a quadratic regression model ($F(2, 140) = 68.22, p < .001$) (See Figure 3), which indicated a floor effect. The range of this floor effect comprised scores of 7 or below, which were scored by 73% of MCI patients and 84% of AD patients. No healthy elderly controls scored within the range of this floor effect. For multiple-choice cued recognition, the explained variance in test performance was not increased by the squared CST scores. Thus, the best fit for the multiple-choice cued recognition data was a linear regression model ($F(1, 142) = 48.32, p < .001$) (See Figure 4), which indicated that multiple-choice cued recognition did not show a floor effect.

Chapter 4

Table 2. Summary of hierarchical regression analysis for CST variables predicting VAT-E paired associate recall, free recall, and multiple-choice cued recognition total scores ($n = 144$).

| VAT-E | Variable | <i>B</i> | <i>SE B</i> | β |
|---|--------------------------|----------|-------------|----------|
| Paired associate recall | Step 1 (linear model) | | | |
| | Constant | -17.16 | 3.2 | |
| | CST | 1.66 | 0.19 | .60*** |
| | Step 2 (quadratic model) | | | |
| | Constant | 17.69 | 11.03 | |
| | CST | -3.03 | 1.44 | -1.08* |
| Free recall | Step 1 (linear model) | | | |
| | Constant | -26.40 | 3.85 | |
| | CST | 2.24 | 0.23 | .64*** |
| | Step 2 (quadratic model) | | | |
| | Constant | 34.50 | 12.79 | |
| | CST | -5.97 | 1.67 | -1.69*** |
| Multiple- choice cued recognition | Step 1 (linear model) | | | |
| | Constant | 1.17 | 1.27 | |
| | CST | 0.53 | 0.08 | .50*** |
| | Step 2 (quadratic model) | | | |
| | Constant | 1.65 | 4.59 | |
| | CST | 0.46 | 0.60 | .44 |
| | CST squared | 0.00 | 0.02 | .06 |

Note. PA: $R^2 = .353$ for Step 1; $\Delta R^2 = .046$ for Step 2 (F change (1, 141) = 10.81, $p < .01$); FR: $R^2 = .405$ for Step 1; $\Delta R^2 = .089$ for Step 2 (F change (1, 140) = 24.58, $p < .001$); MC: $R^2 = .254$ for Step 1; $\Delta R^2 = .0001$ for Step 2 (F change (1, 141) = 0.01, $p = .912$). * $p < .05$; ** $p < .01$; *** $p < .001$. CST = Cognitive Screening Test; VAT-E = Visual Association Test-Extended; n = number of subjects.

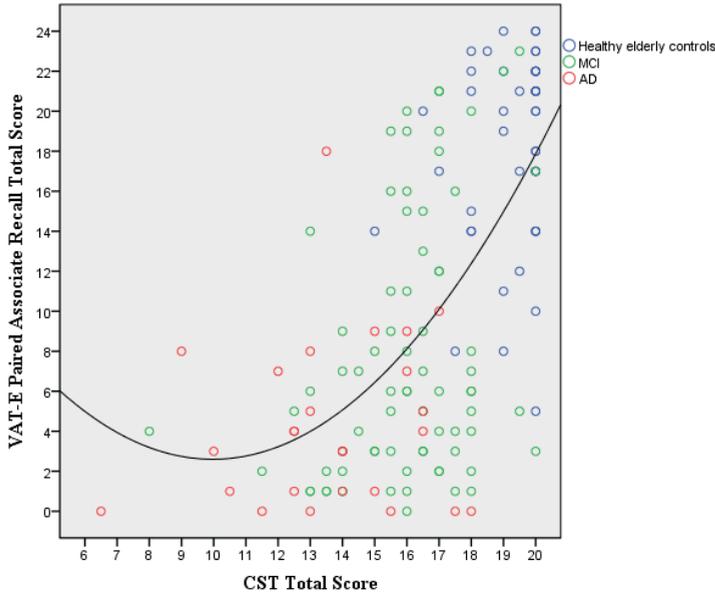


Figure 2. VAT-E paired associate recall scores compared to CST scores ($n = 144$).

Note. VAT-E = Visual Association Test-Extended; CST = Cognitive Screening Test; n = number of subjects; MCI = mild cognitive impairment; AD = Alzheimer’s Disease.

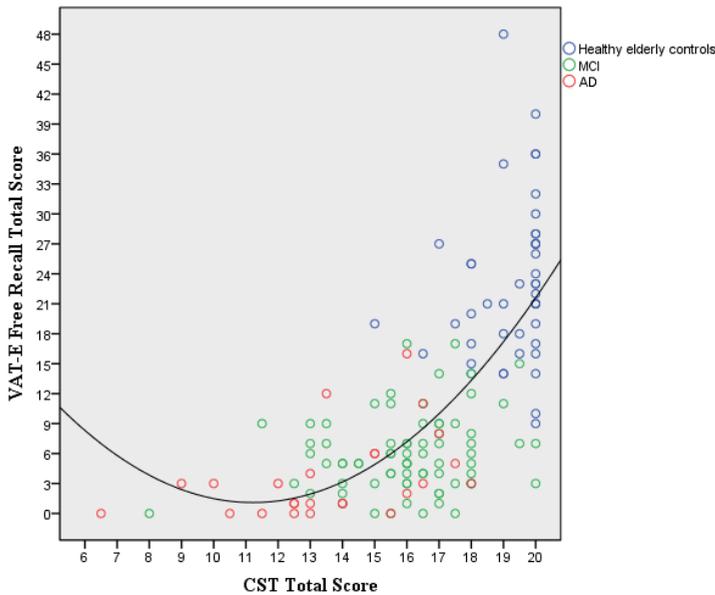


Figure 3. VAT-E free recall scores compared to CST scores ($n = 144$).

Note. VAT-E = Visual Association Test-Extended; CST = Cognitive Screening Test; n = number of subjects; MCI = mild cognitive impairment; AD = Alzheimer’s Disease.

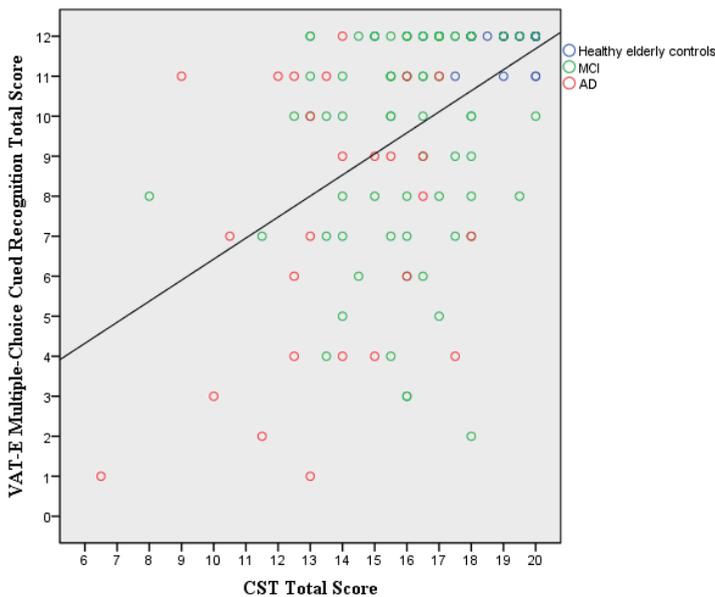


Figure 4. VAT-E multiple choice cued recognition scores compared to CST scores ($n = 144$).

Note. VAT-E = Visual Association Test-Extended; CST = Cognitive Screening Test; n = number of subjects; MCI = mild cognitive impairment; AD = Alzheimer’s Disease.

Discussion

This study aimed to investigate if the VAT-E covers a broad range of memory decline by examining the difficulty levels of several episodic memory measures across a wide range of global cognitive impairment. We compared the outcomes of VAT-E subtests in normal controls, patients with amnesic MCI, and patients with mild AD. Free recall was associated with floor effects in both patient groups, but paired associate recall showed a smaller floor effect, and it was absent in the multiple-choice cued recognition trial. Healthy controls scored well above the floor levels. These results indicate that if strong links are provided between cues and to be remembered material, as is the case with the VAT-E visual associations paradigm, patients with AD remember more as more retrieval support is provided. As is evident by our findings that while free recall of the visual associations was most difficult for patients with MCI or mild AD, cued recall of these visual associations was of intermediate

difficulty, and recognizing newly learned material based on cues was easiest for these patients. Thus, in conjunction with each other, the VAT-E measures can provide insight into the specific degree to which patients within a broad range of cognitive decline can still benefit from retrieval support.

Strong points of our study are that it explores a broad range of episodic memory impairment, from normal cognitive functioning via mild cognitive impairment to the more severe impairment of dementia, and that we used a regression-based method to examine episodic memory in relation to cognitive decline. The use of hierarchical multiple regression analysis enabled us to compare VAT-E episodic memory scores, not only in healthy elderly controls, MCI patients and AD patients, but also in relation to global cognitive impairment. Since we compared paired associate recall, free recall, and multiple-choice cued recognition using the same groups, we consider our findings to be robust.

Our results are in line with previous research. While some test paradigms, such as free recall, are well suited for measuring mild memory impairment, other paradigms, such as paired associate recall or recognition, may be preferred for measuring more severe impairment in AD (Bäckman, Jones, Berger, Laukka, & Small, 2005; Greenaway et al., 2006; Lownes et al., 2008; Tounsi et al., 1999). The VAT-E applies these memory paradigms by providing visual associations in which specific cues are encoded along with associated pictures during incidental learning, and by providing various degrees of retrieval support during testing (Craik & Lockhart, 1972; Tulving & Madigan, 1970). Our results indicate that within the VAT-E testing paradigm, patients with AD remember more as more retrieval support is provided, which is in line with recent findings indicating that previously forgotten information can be retrieved in early AD (Roy et al., 2016).

The visual associations enhance the engagement with the test material, and thus they support the establishment of longer lasting, and stronger memory traces (Craik & Lockhart,

1972; Tulving & Madigan, 1970). This is in line with previous research. First, in contrast to words, pictures are mostly coded visually as well as verbally. Due to this dual coding process (Paivio, 1969; Paivio, 1991), the retrieval of pictures is easier than the retrieval of words (Alley et al., 2009; Hockley, 2008; Paivio et al., 1968; Shepard, 1967). Second, the interaction depicted in the visual association creates a stronger link between the cue and the associated picture than a mere side-by-side presentation of the same pictures. In other words, cued recall of the associations is improved within the incidental learning paradigm compared to cued recall of the separate pictures within an intentional learning paradigm (Lindeboom et al., 2002). Third, the depicted interactions are unexpected, and thus they are elaborated on more than expected interactions (Birngruber, Schröter, & Ulrich, 2014; Ulrich, Nitschke, & Rammsayer, 2006). In addition, providing an occasion of incidental learning gives the test administrator more control over how the subject processes the material, as opposed to intentional learning situations, in which the subject could use a mnemonic strategy that remains hidden to the test administrator (Craik & Lockhart, 1972). That is, as subjects are not explicitly instructed to learn but to perform another task, i.e. picture naming, they will not apply an unknown coding strategy. Thus, encoding will be more equal among subjects. Furthermore, the encoding is facilitated by the dual coding of pictures, the depicted interaction, and the unexpected nature of the association.

A potential limitation of this study is that global cognitive impairment was determined by means of a screening test, i.e. the CST, which might be less sensitive to MCI than to dementia. However, sensitivity is based on cut-off scores, which we did not use in our analysis. Instead, we used the entire measurement range of the CST. Therefore, if this had had an effect on our results, then it would have been evident in a restriction of range for the MCI patients. However, as table 1 shows, variances of CST scores increased considerably from healthy elderly controls to MCI patients to AD patients, i.e. more than 200% and 80%,

respectively. Therefore, we consider the CST to be an adequate tool for determining the range of global cognitive impairment in these groups.

Another limitation is that the clinical neuropsychologist was not blinded to the CST score during the diagnostic process. This was due to the fact that the CST is part of the hospital's standardized neuropsychological examination, and thus it could not be removed from the patient's hospital dossier. However, the CST played only a very small part in the diagnosis of the patients, because it is a screening test, and the diagnostic program of the patients was far more extensive than the CST alone, as it included a comprehensive geriatric assessment, laboratory results, magnetic resonance imaging or computed tomography scanning, and a standardized neuropsychological examination consisting of several tests relating to relevant cognitive functions. Besides, the study protocol did ensure that the clinical neuropsychologist was blinded to the VAT-E scores, which were not used in the diagnostic process, and therefore we consider our results to be robust.

Another possible limitation is that we investigated patients with mild cognitive impairment and mild AD, and thus we did not investigate episodic memory impairment in AD for the whole range of cognitive decline. However, clinicians who want to determine the degree to which patients can still benefit from retrieval support will generally want to do so when the first symptoms of the disease come to the fore, i.e. during the early stages of the disease, which are included in our study.

To determine whether a patient with an episodic memory impairment can still benefit from retrieval support, one has to know not only the norm score, but also the level of difficulty of the measures used. In addition to the norms provided in the manual of the VAT-E (Meyer & de Jonghe, 2017), our analyses show that in relation to the severity of global cognitive impairment, a large proportion of patients performed at floor level on free recall and a smaller proportion of patients performed at floor level on paired associate recall. Only on

the least difficult measure, multiple-choice cued recognition, no patient performed at floor level. Thus, if a patient scores within the impaired range on free recall, this establishes an episodic memory impairment, but it gives no information about the ability of a patient to still benefit from retrieval support. For this, the clinician should also inspect the scores on paired associate recall and multiple-choice cued recognition. If, in addition to free recall, a subject benefits from the retrieval support of paired associate recall and even more so from the retrieval support of multiple-choice cued recognition, this may support health care to be more closely aligned with patients' memory capacities.

If the VAT-E indicates that a patient can still benefit from retrieval support, health care professionals and informal caregivers can be provided with memory tools that improve memory performance in patients' daily functioning. Indeed, previous research showed that learning methods based on cues are effective in mild AD patients (Bier et al., 2008). A patient might be helped to learn to associate activities of daily living with icons, which subsequently could be recognized by that patient in a day planner. Also, health care professionals and informal caregivers could be given tools for communicating with patients, as they could be coached how to align their communication style with the retrieval support the patient needs, by using cues or multiple-choice options in conversation. In addition, we consider the VAT-E to be more suitable for elderly patients than the frequently used word list learning tests, because it is based on visual associations which are encoded by means of incidental learning, making it less taxing for these patients. Also, as opposed to most other visual memory tests, the VAT-E tests memory for pictures of everyday objects, animals, plants, or food instead of abstract designs, which makes it more applicable to patients' daily memory functioning.

Episodic memory tests are needed that enable clinicians to determine the degree to which patients with MCI or AD can still benefit from retrieval support across a broad range of cognitive decline. Our findings show that the VAT-E covers a broad range of episodic

memory decline in patients. In contrast to most other memory tests, the VAT-E uses visual associations and incidental learning to establish longer lasting and stronger memory traces. By subsequently varying retrieval support, we constructed measures of varying levels of difficulty. These varying levels of difficulty enable a more accurate determination of the level of retrieval support that can still benefit patients across a broad range of cognitive decline. Providing patients with the right amount of retrieval support may improve their daily functioning.

Acknowledgments

We thank the department of geriatric medicine of the Northwest Medical Center location Alkmaar, the Netherlands, for enabling us to do our research.

Disclosure Statement

S.R.A.M and J.F.M.J are the authors of the Visual Association Test-Extended.

Funding

This work was supported by Foreest Northwest Medical School [grant number FIO1308].

References

- Alley B. A., Gold, C. A., & Budson, A. E. (2009). The picture superiority effect in patients with Alzheimer's disease and mild cognitive impairment. *Neuropsychologia*, *47*, 595-598.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: Author.
- Bäckman, L., Jones, S., Berger, A., Laukka, E. J., & Small, B. J. (2005). Cognitive

- impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, *19*, 520-531.
- Bier, N., van der Linden, M., Gagnon, L., Desrosiers, J., Adam, S., Louveaux, S., & Saint-Mieux, J. (2008). Face-name association learning in early Alzheimer's disease: A comparison of learning methods and their underlying mechanisms. *Neuropsychological Rehabilitation*, *18*, 343-371.
- Birngruber, T., Schröter, H., & Ulrich, R. (2014). Duration perception of visual and auditory oddball stimuli: Does judgement task moderate the temporal oddball effect? *Attention, Perception, & Psychophysics*, *76*, 814-828.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment. An introduction to tests and measurement* (8th ed., pp. 256, 259). New York: McGraw-Hill.
- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.
- Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*, 2308-2314.
- Dautzenberg, P. L. J., Schmand, B., Vriens, M. T. S., Deelman, B. G., & Hooijer, C. (1991). De validiteit van de cognitieve screening-test en de 'mini-mental state examination' bij een groep oudere ziekenhuispatiënten [Study of the validity of the cognitive screening test and the mini-mental state examination in an older hospital population]. *Nederlands Tijdschrift voor Geneeskunde*, *135*, 850-855.
- De Graaf, A. & Deelman, B.G. (1991). *Cognitive Screening Test. Manual for administration and scoring*. Lisse: Swets en Zeitlinger bv.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning*

- Test (CVLT). Adult Version. Manual.* San Antonio, TX: Psychological Corporation.
- Dierckx, E., Engelborghs, S., de Raedt, R., van Buggenhout, M., de Deyn, P. P., Verté, D., & Ponjaert-Kristoffersen, I. (2009). Verbal cued recall as a predictor of conversion to Alzheimer's disease in mild cognitive impairment. *International Journal of Geriatric Psychiatry, 24*, 1094-1100.
- Diesfeldt, H., Prins, M., & Lauret, G. (2017). De visuele associatietest (VAT) als instrument voor de ouderenpsycholoog [The Visual Association Test to study episodic memory in clinical geriatric psychology]. *Tijdschrift voor Gerontologie en Geriatrie*. Advanced online publication. doi:10.1007/s12439-017-0231-7.
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ...Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology, 6*, 734-746.
- Folstein, M. F., Folstein, S. F., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.
- Fox, L. S., Olin, J. T., Erblich, J., Ippen, C. G., & Schneider, L. S. (1998). Severity of cognitive impairment in Alzheimer's disease affects list learning using the California Verbal Learning Test (CVLT). *International Journal of Geriatric Psychiatry, 13*, 544-549.
- Fuchs, A., Wiese, B., Altiner, A., Wollny, A., & Penzek, M. (2011). Cued recall and other cognitive tasks to facilitate dementia recognition in primary care. *Journal of American Geriatric Society, 60*, 130-135.
- Greenaway, M. C., Lacritz, L. H., Binegar, D., Weiner, M. F., Lipton, A., & Cullum, C. M. (2006). Patterns of verbal memory performance in mild cognitive impairment, Alzheimer disease, and normal aging. *Cognitive & Behavioral Neurology, 19*, 79-84.

- Hammes, J. G. W. (1971). *De Stroop Kleur-Woord Test. Handleiding*. [The Stroop Colour-Word Test. Manual]. Amsterdam: Pearson Assessment and Information B.V.
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, *36*, 1351-1359.
- Hughes, C. P., Berg, L., Danziger, W. L., Coben, L.A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, *140*, 566-572.
- Jorm A. F. & Jacomb P. A. (1989). The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): Socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine*, *19*, 1015-1022.
- Kahn, R. L., Goldfarb, A. I., Pollack, M., & Peck, A. (1960). Brief objective measures for the determination of mental status in the aged. *The American Journal of Psychiatry*, *117*, 326-328.
- Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychological function. *JAMA*, *185*, 914-919.
- Lezak, M. D., Howieson, D. B., Bigler, E., & Tranel, T. (2012). *Neuropsychological Assessment* (5th ed., pp. 260). New York, NY: Oxford University Press.
- Lindeboom, J. & Jonker, C. (1989). *Amsterdam Dementie-Screening Test*. [Amsterdam Dementia-Screening Test]. Amsterdam: Pearson Assessment and Information B.V.
- Lindeboom, J. & Schmand, B. (2003). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association

- test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73, 126-133.
- Lowndes, G. J., Saling, M. M., Ames, D., Chiu, E., Gonzalez, L. M., & Savage, G. R. (2008). Recall and recognition of verbal paired associates in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, 14, 591-600.
- Lawton, M. P. & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179-186.
- Luteijn, F. & Barelds, D. P. H. (2004). *GIT 2: Groninger Intelligentie Test 2*. [GIT 2: Groninger Intelligence Test 2]. Amsterdam: Harcourt Assessment B.V.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual]. Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A., de Jonghe, Schmand B., & Ponds, R. W. H. M. (2017). The Visual Association Test-Extended: A cross-sectional study of the performance validity measures. *The Clinical Neuropsychologist*, 31, 789-813.
- Meyer, S. R. A., Spaan, P. E. J., Boelaarts, L., Ponds, R. W. H. M., Schmand, B., & de Jonghe, J. F. M. (2016). Visual associations cued recall A paradigm for measuring episodic memory decline in Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, 23, 566-577.
- Morris, J. C. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43, 2412-2414.
- Osterreith, P. A. (1944). *Le test de copie d'une figure complex* [The Complex Figure Copy Test]. *The Clinical Neuropsychologist*, 7, 9-15.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 74, 241-263.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of*

- Psychology*, 45, 255-287.
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11, 137-138.
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183-194.
- Petersen, R. C., Doody, R., Kurz, A. Mohs, R. C., Morris, J.C., Rabins, P. V., ... Winbald, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985-1992.
- Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society*, 23, 433-441.
- PsychCorp. (2009). *Wechsler Memory Scale-Fourth Edition (WMS-IV) technical and interpretive manual*. San Antonio, TX: Pearson.
- Reesink, F. E., Lemstra, A. W., van Dijk, K. D., Berendse, H. W., van de Berg, W. D. J. Klein, M., ... van de Flier, W. M. (2010). CSF α -synuclein does not discriminate dementia with Lewy bodies from Alzheimer's disease. *Journal of Alzheimer's Disease*, 22, 87-95.
- Reitan, R. M. & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery*. Tucson, AZ: Neuropsychological Press.
- Rey, A. (1941). *L'examen psychologique dans les cas d'encephalopathie traumatique* [Psychological examination of traumatic encephalopathy]. *Archives de Psychologie*, 28, 286-340.
- Rey, A. (1964). *L'examen Clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Roth, M., Huppert, F. A., Mountjoy, C. Q., & Tym, E. (1998). *The Cambridge Examination*

- of Mental Disorders of the Elderly Revised: CAMDEX-R*. Cambridge: Cambridge University Press.
- Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T.J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature*, *531*, 508-512.
- Saan, R. J., & Deelman, B. G. (1986). *15-Woorden Test A en B* [15-Word Test A and B]. In Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (Eds.) (2012). *Handboek neuropsychologische diagnostiek* [Handbook neuropsychological assessment] (2nd ed., p.p. 267-282). Amsterdam: Pearson Assessment and Information B.V.
- Schmand, B., Deelman, B. G., Hooijer, Ch., Jonker, C., & Lindeboom, J. (1996). De itemreeks van de cognitieve screening test vergeleken met die van de mini-mental state examination [The items of the cognitive screening test compared to those of the mini-mental state examination]. *Tijdschrift voor Gerontologie en Geriatrie*, *27*, 29-33.
- Schmidt, M. (1996). *Rey Auditory Verbal Learning Test. A handbook*. Los Angeles: Western Psychological Services.
- Sheik, J. I. & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*, 165-173.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156-163.
- Sivan, A. B. (1992). *Benton Visual Retention Test* (5th ed.). San Antonio, TX: PsychoCorp/Pearson.
- Tounsi, H., Deweer, B., Ergis, A., van der Linden, M., Pillon, B., Michon, A., & Dubois, B. (1999). Sensitivity to semantic cueing: An index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Disease and Associated Disorders*, *13*, 38-46.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of*

- Psychology*, 21, 437-484.
- Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research*, 70, 77-87.
- UNESCO (2006). *International Standard Classification of Education: ISCED 1997 (re-edition)*. Montreal: UNESCO-UIS.
- Van der Vlies, A. E., Verwey, N. A., Bouwman, F. H., Blankenstein, M. A., Klein, M., Scheltens, P., & van der Flier, W. M. (2009). CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology*, 72, 1056-1061.
- Van Toutert, M., Diesfeldt, H., & Hoek, D. (2016). De cognitieve screening test (cst) bij normale cognitieve veroudering en dementie: Drie varianten en bruikbaarheid voor de klinische praktijk [The cognitive screening test (cst) for normal cognitive functioning and dementia: Three subsets and usefulness for clinical practice]. *Tijdschrift voor Neuropsychologie*, 3, 274-292.
- Verhage, F. (1964). Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar [Intelligence and age: Research within Dutch participants aged from 12 to 77 years old]. Dissertation. Assen: Van Gorcum.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, 19, 87-95.
- Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St. Edmunds, U.K.: Thames Valley Test Company.

5

Chapter 5

TESTING EPISODIC MEMORY IN ELDERLY SUBJECTS: NOT AS SIMPLE AS IT LOOKS

Leo Boelaarts, Sascha R.A. Meyer, Philip Scheltens, and Jos F.M. de Jonghe

Dementia and Geriatric Cognitive Disorders, 2019; 9: 207-216

Abstract

The available episodic memory tests are not specifically constructed to examine older subjects. Their use in outpatient memory clinics may result in aborted test administration. We used a strict adherence to a test protocol, i.e. intensified test guidance, in cognitively healthy, amnesic mild cognitive impairment (aMCI), and Alzheimer's disease dementia subjects to assess the possibility of preventing this. This is a cross-sectional study in memory outpatient subjects with a mean age of 74.5 years. Primary study outcomes were: number of missing values and test results in the Visual Association Test (VAT) and the 15 Word Test (15WT). A strict adherence to a test protocol, i.e. intensified test guidance, resulted in a 10-fold decrease in the number of missing values in the VAT. For the 15WT this could not be realized mostly because the test was deemed too demanding for 1 in 6 patients. This study is one of the few examining the applicability of well-known episodic memory tests in older subjects. A strict adherence to a test protocol, i.e. intensified test guidance, reduced the number of missing values. Floor effects were stronger for the 15WT than for the VAT. Results favor the use of the VAT in senior subjects and show the unsuitability of the 15WT in this group.

Introduction

Elderly subjects presenting with cognitive problems in an outpatient memory clinic are usually assessed using a standard battery of cognitive tests. Results of episodic memory tests may indicate incipient Alzheimer's disease (AD) (Dubois et al., 2007; Hort et al., 2010; Knopman et al., 2001; McKhann et al., 2011; Sorbi et al., 2012).

Few if any episodic memory tests have been specifically developed for elderly subjects. Among the available episodic memory tests, the Rey Auditory Verbal Learning Test (RAVLT) is commonly used (Rey, 1958). The Dutch adaptation of the RAVLT is the 15 Word Test (15WT; Saan & Deelman, 1986), which may be strenuous and easily tiring even for cognitively healthy subjects (Van der Elst, van Boxtel, van Breukelen, & Jollis, 2005), thereby showing strong floor effects. Age-related changes in cognition are mainly caused by a reduction of the speed of perception and reasoning and especially affect working memory capacity (Bors & Forrin, 1995). A measurable decrement in these functions can be found as a linear function of increasing age going from 18 to 50 years. This effect accelerates after the age of 50 years (Verhaeghen & Salthouse, 1997; Bopp & Verhaeghen, 2005).

Meyer et al. (2016) demonstrated that the less taxing Visual Association Test (VAT) (Lindeboom & Schmand, 2003; Lindeboom, Schmand, Meyer, & de Jonghe, 2014; Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002; Meyer, de Jonghe, Schmand, & Ponds, 2017; Meyer et al., 2016) is a sensitive and specific measure of episodic memory function and at the same time shows less floor effects in older subjects compared to the 15WT. Patients who refuse to take a test or who stop the administration of a test due to fatigue because of an advanced age may generate missing test values, potentially leading to result bias.

As far as we know, no studies have reported on the number of elderly patients discontinuing the 15WT or the VAT, or on ways to decrease the number of missing values.

We hypothesized that focusing on ways in the administration of the tests to minimize the number of missing values, a strict adherence to a test protocol, i.e. intensified test guidance, is feasible and will also increase test results. In addition, we studied the effect of age on the number of missing values and test results.

Methods

Study design

This is a cross-sectional study examining the number of missing values in a strict adherence to a test protocol, i.e. intensified test guidance (study A; the experimental condition) and in routine administration of episodic memory tests (study B; the control condition).

Patients

Patients were recruited from an outpatient memory clinic population and assessed in a diagnostic program comprised of taking a medical history, a physical examination, a standardized neuropsychological test battery, laboratory tests, ECG, and MRI or CT. Patients were randomly selected to take the VAT or a parallel version of the VAT and took an extra VAT and 15WT subtest; in the present study this comprised the study A group. A smaller control group, study B, was created, comprised of patients who were assessed using the usual test protocol. The patient characteristics and the neuropsychological test battery were described earlier (Boelaarts, Scheltens, & de Jonghe, 2016; Meyer et al., 2016). For the present study, results of patients not able to read letters of 2 mm in height from a comfortable reading distance and patients with a second diagnosis potentially influencing the test results, e.g., depression or anxiety disorder, were excluded.

We used the original VAT version in study A, resulting in roughly half the number of patients in the delayed paired associate recall and recognition condition in comparison to the immediate paired associate recall condition.

The data used in the present study were of subjects who had received 1 of the following 3 diagnoses: no cognitive disorder (NCD), amnesic mild cognitive impairment (aMCI) according to the Petersen criteria (Petersen et al., 1999), and AD dementia (ADD) using the NINDS-ADRDA criteria (McKahn et al., 2011).

Administration of the VAT and the 15WT

The 15WT is in the test battery used to prevent ceiling effects and the VAT is used to prevent floor effects.

As the goal of study A was to validate the VAT, it was essential to have also data of more severely cognitively impaired patients. Psychotechnicians were specifically instructed in an attempt to minimize the number of missing values but the administration of the tests per se was identical in both protocols and test results were not expected to be different beforehand. In the Appendix, these instructions and the usual instructions for psychotechnicians are shown.

The psychotechnician reported in categories “not administered”, indicating that the patient refused to take the test or the test leader assessing the patient determined that the patient was incapable of taking the test for any reason; “stopped during administration”, denoting refusal of the patient to continue the test; “missing values due to other causes”. This information was not available for study B patients; “missing values” here indicated all-cause missing values.

The Dutch Cognitive Screening Test is administered in our test battery. For this study the results of this test were converted into Mini-Mental State Examination (MMSE) scores. This test and the conversion operation have been well validated (Schmand, Deelman, Hooijer, Jonker, & Lindeboom, 1996).

Statistical analysis

Data were analyzed using the SPSS version 20 statistical package for Windows and the MEDCALC® statistical package. Results are expressed as numbers (%) or means \pm SD. One-way analysis of variance (ANOVA) was used to study the effect of diagnostic group membership on age and MMSE, as well as the relation between age and membership of a diagnostic group and education level. The χ^2 statistic was used to study the sex ratio and diagnosis. The relation between age and the number of missing values was studied using the Fisher exact test. To compare means of test results the paired samples or independent samples t test statistic was used and in case of a small number of cases the nonparametric Kruskal-Wallis test was used. We employed a linear regression analysis to examine the effect of age on the results of the VAT and the 15WT in the 3 diagnostic groups. $p < 0.05$ was considered a statistically significant effect.

Results

Study A (Tables 1 and 2)

Clinical features, demographics, and mean test results of the diagnostic groups

The participating patients may be considered to represent a typical sample of outpatients in a geriatric department; 50% of the patients used more than 3 drugs, and 10–20% of the patients had a history of heart failure, atrial fibrillation, coronary pathology, cerebrovascular and peripheral artery disease, hypertension, hypercholesterolemia, or diabetes.

aMCI and ADD patients were older and had lower MMSE scores (ANOVA, $p = 0.000$ and $p = 0.000$, respectively). Seventy-one percent of the NCD subjects were female, in both of the other diagnostic groups gender was more evenly distributed (χ^2 , $p = 0.018$). Fourteen percent of the subjects had a low education level, 59% had a medium level, and 27% had a

higher education. In the univariate analysis of variance, age was a significant factor for diagnostic group membership but education level was not.

Immediate paired associate recall as measured with the VAT overtaxed the ability to store and recall information in aMCI and ADD patients as none of these patients scored above the cut-off value of 9. The results in the delayed paired associate recall trial were comparable to the immediate paired associate recall trial results; all aMCI and ADD patients scored below the delayed recall cut-off value of 4. Remarkably, aMCI and ADD patients significantly improved on the VAT recognition trial compared to their delayed paired associate recall scores (t test, $p = 0.000$). NCD subjects outperformed aMCI patients on 15WT immediate free recall, and aMCI patients did better than ADD patients. aMCI and ADD patients did equally poor on 15WT delayed free recall. Both groups performed much better on 15WT recognition compared to immediate and delayed free recall (paired t test, $p = 0.000$).

Missing values, effect of age on number of missing values, and mean test results

None of the patients refused to take the VAT and all of the patients completed this test. Comparing all 3 VAT subtests, we found that maximally 2.2% of all cases had missing values for miscellaneous reasons.

In the 15WT, we found for all cases maximally 18.7% missing values due to all possible causes. In maximally 16.5% of all cases this test was not administered, in all but 1 case this occurred in the ADD group. The test was aborted in 3 ADD subjects, comprising 1.6% of all cases. Missing values were reported due to other reasons in 1 patient, comprising 0.5% of all cases.

Age was not associated with the number of missing values for the VAT. A strong age effect was found for the 15WT number of missing values.

Within each diagnostic group, age was not associated with any of the VAT subtest results. Age was significantly associated with all three 15WT subtest scores in NCD subjects (immediate free recall, $p = 0.003$, $R^2 = 12.8\%$; delayed free recall, $p = 0.001$, $R^2 = 15\%$; and recognition, $p = 0.001$, $R^2 = 15\%$, respectively). No significant age effects on 15WT scores were found in the aMCI group or in the ADD group.

Table 1. Demographics and VAT test results for all patients and diagnosis groups (study A)

| | All | NCD | aMCI | ADD |
|--|----------|----------|----------|----------|
| Subjects | 182 | 68 | 46 | 68 |
| Age, years | 75.5±8.9 | 70.0±9.1 | 77.0±7.7 | 79.9±6.2 |
| Female (%) | 105 (58) | 48 (71) | 25 (54) | 32 (47) |
| MMSE score | 24±5.1 | 28±0.96 | 25±3.1 | 20±5.5 |
| <i>VAT immediate paired associate recall sum results of 2 trials (maximum score: 12)</i> | | | | |
| Subjects | 178 | 68 | 44 | 66 |
| Not administered | 0 | 0 | 0 | 0 |
| Stopped during testing | 0 | 0 | 0 | 0 |
| Missing values | 4 | 0 | 2 | 2 |
| Mean | 7.2±4.7 | 11.2±1.3 | 7.3±3.8 | 2.8±3.4 |
| <i>VAT delayed paired associate recall (maximum score: 6)</i> | | | | |
| Subjects | 96 | 36 | 30 | 30 |
| Not administered | 0 | 0 | 0 | 0 |
| Stopped during testing | 0 | 0 | 0 | 0 |
| Missing values | 1 | 0 | 1 | 0 |
| Mean | 3.8±2.5 | 5.8±0.45 | 3.8±2.1 | 1.4±2.0 |
| <i>VAT delayed multiple-choice cued recognition (maximum score: 6)</i> | | | | |
| Subjects | 96 | 36 | 30 | 30 |
| Not administered | 0 | 0 | 0 | 0 |
| Stopped during testing | 0 | 0 | 0 | 0 |
| Missing values | 1 | 0 | 1 | 0 |
| Mean | 4.8±2.0 | 6.0±0.17 | 5.7±0.71 | 2.5±2.2 |

Notes: Values are presented as numbers or means ± SD; VAT = Visual Association Test; NCD = no cognitive disorder; aMCI = amnesic mild cognitive impairment; ADD = Alzheimer's disease dementia; female (number and proportion of females); MMSE = Mini-Mental State Examination.

Chapter 5

Table 2. 15WT results for all of the patients and diagnosis groups (study A)

| | All | NCD | aMCI | ADD |
|---|-----------|-----------|----------|----------|
| Subjects | 182 | 68 | 46 | 68 |
| <i>15WT immediate free recall sum result (maximum score: 75)</i> | | | | |
| Subjects | 152 | 67 | 46 | 39 |
| Not administered | 29 | 1 | 0 | 28 |
| Stopped during testing | 1 | 0 | 0 | 1 |
| Mean | 29.3±13.9 | 40.9±11.1 | 22.7±7.1 | 17.3±7.8 |
| <i>15WT delayed free recall sum result (maximum score: 15)</i> | | | | |
| Subjects | 151 | 66 | 46 | 39 |
| Not administered | 29 | 1 | 0 | 28 |
| Stopped during testing | 1 | 0 | 0 | 1 |
| Missing values | 1 | 1 | 0 | 0 |
| Mean | 4.8±4.7 | 8.5±3.7 | 1.9±1.7 | 1.8±4.5 |
| <i>15WT delayed recognition total result of falsely recognized items (maximum number of errors: 30)</i> | | | | |
| Subjects | 148 | 66 | 46 | 36 |
| Not administered | 30 | 1 | 0 | 29 |
| Stopped during testing | 3 | 0 | 0 | 3 |
| Missing values | 1 | 1 | 0 | 0 |
| Mean faults | 5.3±4.7 | 1.5±2.1 | 7.5±4.2 | 9.4±3.3 |

Notes: Values are presented as numbers or means ± SD; 15WT = Dutch version of the Rey Auditory Verbal Learning Test; NCD = no cognitive disorder; aMCI = amnesic mild cognitive impairment; ADD = Alzheimer's disease dementia.

Study B (Table 3)

Clinical features, demographics, and mean test results of the diagnostic groups

The education levels and patient characteristics were comparable to those of the study A group; aMCI and ADD patients were older and had lower MMSE scores. In this study 71% of NCD subjects were male, with a higher percentage of women in the other diagnostic groups (χ^2 ns). Mean test results and SD were comparable to those of study A.

Missing values, effect of age on number of missing values, and mean test results

In both VAT conditions in all cases maximally 17.8% missing values were reported. In the NCD group most missing values were reported (maximally 28.6%), in comparison to the aMCI subjects (6.7%) and the ADD group (18.8%).

For the 15WT the maximum percentage of missing values for all cases was 20%. For this test the percentage of missing values was lowest in the NCD group (7.1%) in comparison to aMCI patients (13.3%) and the ADD group (37.5%).

No age effect was found in the number of missing values in either test (χ^2 0.811). The number of patients who had missing values either in the VAT or in the 15WT amounted to 23 (13 younger and 10 older than the mean age of 73.5 years).

Within all diagnostic groups no age effect was found on any mean test result, with one exception, i.e. the control study B group in the 15WT delayed recognition condition ($p = 0.037$, $R^2 = 34\%$).

Table 3. Demographics and VAT and 15WT test results for all patients and diagnosis groups (study B)

| | All | NCD | aMCI | ADD |
|---|----------|----------|----------|----------|
| Subjects | 45 | 14 | 15 | 16 |
| Age, years | 73.5±9.0 | 65.6±8.3 | 75.2±6.7 | 78.8±6.8 |
| Female (%) | 23 (51) | 4 (29) | 10 (67) | 9 (56) |
| MMSE score | 24±4.5 | 28±1.5 | 25±2.5 | 20±4.1 |
| <i>VAT immediate paired associate recall sum of 2 trials (maximum score: 12)</i> | | | | |
| Subjects | 39 | 11 | 15 | 13 |
| Missing values | 6 | 3 | 0 | 3 |
| Mean | 7.1±4.1 | 10.8±2.4 | 6.3±3.5 | 4.9±3.8 |
| <i>VAT delayed paired associate recall sum result of 2 trials (maximum score: 12)</i> | | | | |
| Subjects | 37 | 10 | 14 | 13 |
| Missing values | 8 | 4 | 1 | 3 |
| Mean | 4.0±3.0 | 6.5±2.0 | 3.8±3.2 | 2.2±1.9 |
| <i>15WT delayed free recall sum result (maximum score: 15)</i> | | | | |
| Subjects | 37 | 13 | 13 | 11 |
| Missing values | 8 | 1 | 2 | 5 |
| Mean | 3.5±3.6 | 6.8±4.0 | 2.2±1.9 | 1.2±1.7 |
| <i>15WT delayed recognition total result of falsely recognized items (maximum number of errors: 30)</i> | | | | |
| Subjects | 36 | 13 | 13 | 10 |
| Missing values | 9 | 1 | 2 | 6 |
| Mean faults | 7.1±5.5 | 3.5±3.7 | 7.2±3.3 | 11.9±6.2 |

Notes: Values are presented as numbers or means ± SD; VAT = Visual Association Test; 15WT = Dutch version of the Rey Auditory Verbal Learning Test; NCD = no cognitive disorder; aMCI = amnesic mild cognitive impairment; ADD = Alzheimer's disease dementia; female (number and proportion of females); female (number and proportion of females); MMSE = Mini-Mental State Examination.

Comparison of mean test results within diagnostic groups between study A and study B

Data for comparison were available for the VAT immediate paired associate recall and 15WT delayed free recall and recognition conditions. The only significant differences were found in ADD patients who scored significantly higher in study B in the VAT immediate paired

associate recall (t test, $p = 0.03$), and in the 15WT delayed recognition condition the control group in study A scored significantly better than in study B (t test, $p = 0.008$).

Discussion

In 2 studies in healthy, aMCI, and ADD subjects we found a profound effect of strict adherence to a protocol, i.e. intensified test guidance, on the number of missing values. Age did not affect the number of missing values or the mean test results, with the exception of younger subjects (they performed best in the 15WT).

For the VAT the all-cause number of missing values in study A was maximally 2.2% and it was 17.8% in study B. Remarkably, the highest number of missing values in study B was found in NCD subjects, which could be caused by a ceiling effect, e.g., growing bored or experiencing the test as childish. This effect could, however, be cancelled out in this group by the study A protocol. In the study B ADD group, even in the less demanding VAT, we found a high number of 18.8% of missing values. In study A this was maximally 3.0%.

For the 15WT the all-cause numbers of missing values were similar in both studies. This was caused by the high percentage of missing values in ADD subjects. Given the low percentage of ADD patients who stopped the test during administration, the high percentage of “not administered” annotations in study A represented the test leader’s decision not to start the test. This may illustrate the unsuitability of this test for use in ADD subjects. The success of the study A test protocol is also illustrated by the low number of missing values in the NCD and aMCI subjects in comparison to study B.

Based on our findings we conclude that the number of missing values for test results of senior NCD, aMCI, and ADD subjects may be reduced if tests are not too strenuous, burdening their limited memory capacity too much (Bopp & Verhaeghen, 2005; Bors & Forrin, 1995; Van der Elst et al., 2005; Verhaeghen & Salthouse, 1997). These results also

highlight the importance of specifically training psychotechnicians to test this group of senior patients.

The second conclusion is that strict adherence to a test protocol, i.e. intensified test guidance did not generally improve test results. The apparent contradiction of the significantly better VAT immediate paired associate recall test results for the ADD group in study B may well be caused by the success of the study A test protocol. In study A also more demented patients completed their tests, which probably led to lower mean test results for this group in comparison to the ADD study B group. In accordance with this notion, the controls in study A performed better in the 15WT delayed recognition subtest; a strict adherence to the test protocol can help subjects to achieve better.

At the individual patient level, the effect of the attempts to decrease the number of missing values will be less important than the profile of test results and test behavior. However, when categorizing subjects into diagnostic groups based on test results or examining efficacy of new therapies in different diagnostic groups, the number of missing values and better test results may be critically important.

The third conclusion is that age did not affect the number of missing values; for the VAT this was probably due to the small number of missing values. For the 15WT this result may be confounded by the older, ADD subjects. Importantly we did not compare subjects under 50 years of age or over 80 years; the mean age was 73.5 years, meaning that all of the subjects may be characterized as senior.

The fourth conclusion is that we did not find an age effect on the mean test results, with the exception of the group of controls. The VAT is designed to minimize floor effects and thereby the effect of age. By contrast, in accordance with the literature, younger subjects in this study performed better on the 15WT because this test draws heavily on working

memory capacity and speed of processing (Van der Elst et al., 2005). Also the effect of the diagnosis of aMCI or ADD probably masks any age effect with this number of subjects.

Most of the available episodic memory tests such as the RAVLT have been developed with the aim of assessing younger subjects suffering of varying ailments. Spaan, Raaijmakers, and Jonker (2003) provided an overview of the available episodic memory tests and concluded that episodic memory cannot be adequately tested when a neurodegenerative disorder is suspected due to the fact that these tests were not developed to just measure this. Cerami (2017) examined the extent to which episodic memory tests can be used as biomarkers. Rigorous criteria used to evaluate the methodology in pharmaceutical research were applied to evaluate the robustness of episodic memory tests in this respect. All tests failed to fulfil these criteria.

The aim of a test battery is to obtain reliable data in a standardized way, which means that repeated testing will yield identical results. This advantage is balanced by a decrement in ecological validity. An extensive assessment of cognitive function using demanding tests may produce a distorted representation of the actual cognitive ability. Even worse, submitting senior patients to demanding tests may discourage them, leading to interference with the participation observation by the psychotechnician and loss, in this way, of important information to make a clinical assessment of behavior and cognition.

If we were to generalize the 15WT results of this study, it could mean that, even before starting the test administration in the usual way in aMCI and ADD senior subjects, a priori chances are that in 15–45% of the cases no test results will be generated. Even when using a strict adherence to a test protocol, i.e. intensified test guidance, the a priori chance of missing out on 15WT test results may still be around 40% in ADD subjects. This underlines the unsuitability of this test to study episodic memory especially in these patients.

Notably, we found that in ADD and aMCI patients recognition is much better than free

recall. Perhaps this is due to a milder stage of AD in our patients causing a less compromised learning ability compared to patients described in the older literature who already suffered from a more advanced stage of the disease.

aMCI and ADD patients showed comparable results in the VAT and delayed 15WT free recall condition. Previous research has shown an improved memory performance by these subjects when retrieval support by means of cueing or recognition is provided (Bäckman, Jones, Berger, Laukka, & Small, 2015). This illustrates a continuum of episodic memory test performance that is mildly to severely impaired, going from the aMCI to the ADD stage (Greenaway et al., 2006).

The uneven gender distribution across the diagnostic groups, especially in the control group, is coincidental as all patients were selected from the same outpatient clinic. Normative data show that women perform somewhat better on verbal learning tests such as the 15WT (Van der Elst et al., 2005). However, the diagnosis will have a more profound effect on test results and will probably dwarf these gender effects.

To summarize, the main conclusions of this study are the positive effect of a strict adherence to a test protocol, i.e. intensified test guidance, and the unsuitability of the 15WT to test episodic memory in elderly aMCI and ADD subjects.

Our results underline the need to develop more ecologically valid tests for this group. In the meantime, we would like to propose the use of the VAT as an episodic memory test in these subjects.

Statement of ethics

The study protocol was approved by the local ethics committee. All of the patients gave written informed consent.

Disclosure statement

S. Meyer and J. de Jonghe are the authors of the parallel and extended versions of the VAT.

Funding sources

No funding was received.

Author contributions

L. Boelaerts designed this study, analyzed the data, and wrote this article. S. Meyer carried out this study, built the data base and collected the data, and assisted in writing this article.

Ph. Scheltens assisted in designing this study and in writing this article. J. de Jonghe assisted in designing this study, analyzing the data, and writing this article.

Appendix: Instructions for Psychotechnicians

Intensified test guidance (study A)

The awareness of the psychotechnicians regarding the importance of the full completion of the VAT and the 15WT was raised by the research leader of study A by explaining the study goals of study A but also the potential benefit for the patient completing all of the tests and thereby gaining a reliable assessment of the cognitive functions of the subject.

Subsequently the psychotechnicians were trained to extra motivate and support patients. It was explained that the assessment of the psychotechnician of the ability of a patient to take a test remains subjective and that patients may very well be capable of taking a test or achieving better results if they are properly supported and encouraged to give it a try and do the best they can.

Instructions to the patient at the start of the test administration could be as follows: “it is important to give your best to better understand your complaints and the possible causes for

these problems.”

The psychotechnicians were instructed to focus on complimenting patients during administration for correct answers and for their effort in achieving as good as possible results, for example, by saying: “that answer is correct,” and “well done for giving your best.” Also, when experiencing a loss of focus, patients were instructed as follows: “please try again,” “well done, this answer is indeed correct,” “well done, this answer is also correct,” and “take your time, no need to rush.”

Each session was evaluated afterward by the psychotechnician and the research leader to potentially minimize further missing values and maximize test achievement.

Routine test administration (study B)

Routine clinical cognitive testing requires that psychotechnicians use standard test instructions. They are instructed to use a standard test protocol and get objective information on multiple cognitive domains. If a patient finds it difficult to understand test instructions or if he or she fails during the try-out phase, this test will not be completed.

Test results will be discussed by the psychotechnicians and the senior neuropsychologist in multidisciplinary consensus meetings and they will be presented by the physician in the visit when all of the results of the diagnostic program are presented to the patient and informant.

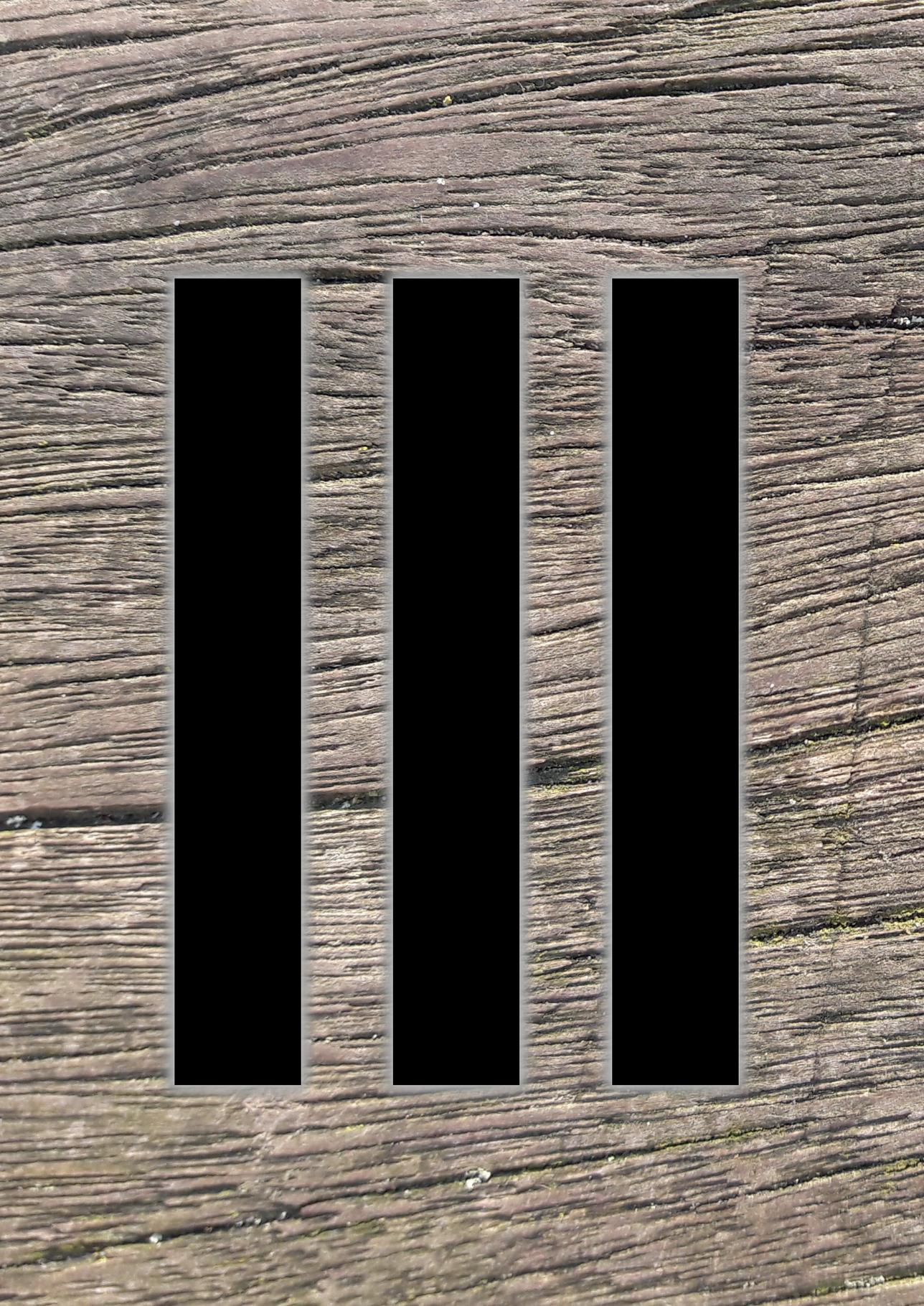
References

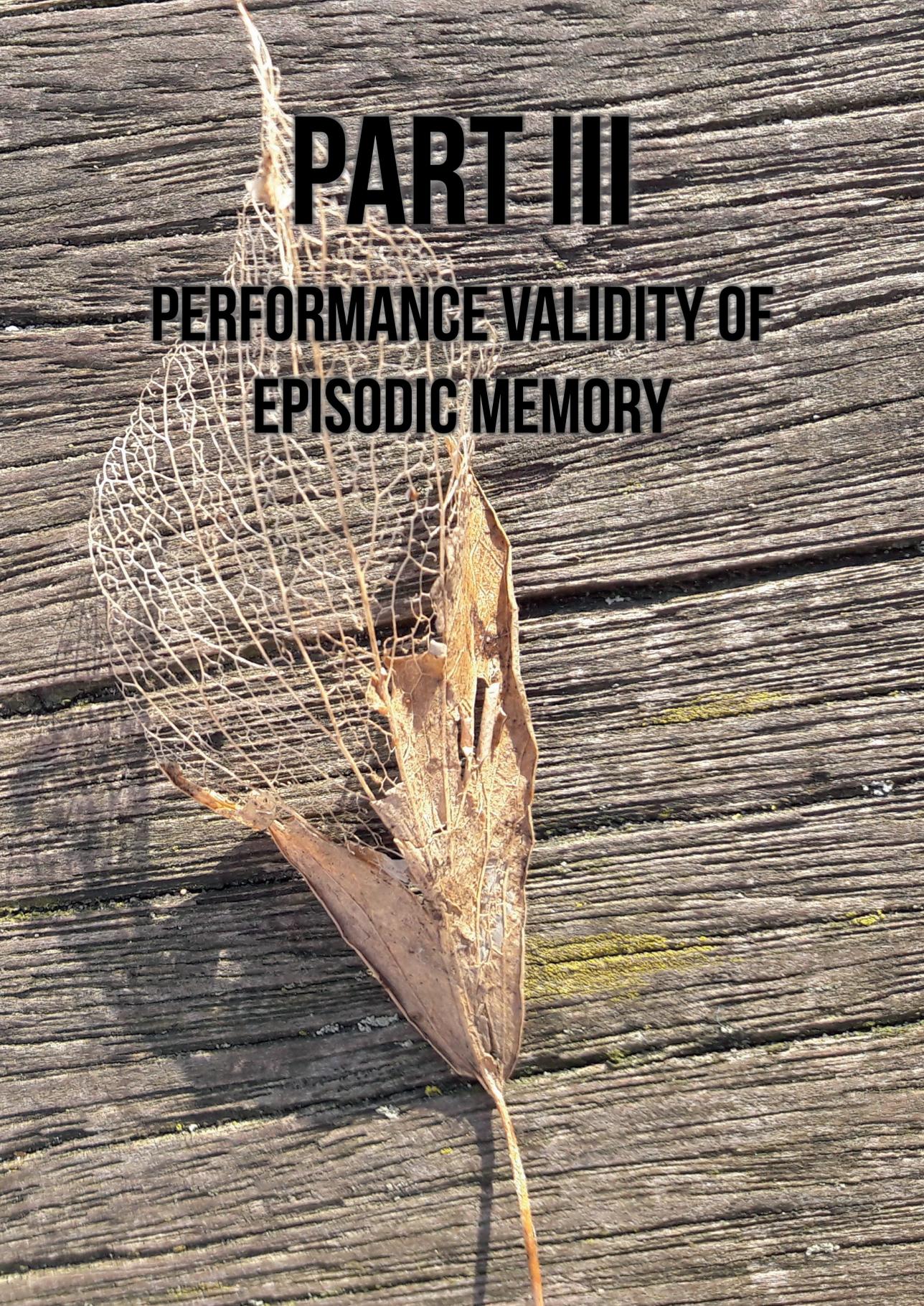
- Bäckman, L., Jones, S., Berger, A., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer’s disease: A meta-analysis. *Neuropsychology, 19*, 520-531.
- Boelaarts, L., Scheltens, P., & de Jonghe J. (2016). Does MRI increase the diagnostic

- confidence of physicians in an outpatient memory clinic. *Dementia and Geriatric Cognitive Disorders*, 6, 242–251.
- Bopp, K. L. & Verhaeghen, P. (2005). Aging and verbal memory span: a meta-analysis. *Journal of Gerontology*, 60B, 223–233.
- Bors, D. A. & Forrin, B. (1995). Age, speed of information processing, recall, and fluid intelligence. *Intelligence*, 20, 229–248.
- Cerami, C., Dubois, B., Boccardi, M., Monsch, A. U., Demonet, J. F., & Cappa, S. F. (2017). Geneva Task Force for the Roadmap of Alzheimer’s Biomarkers. Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer’s disease in the context of a structured 5-phase development framework. *Neurobiology of Aging*, 52, 153–166.
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ...Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer’s disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, 6, 734-746.
- Greenaway, M. C., Lacritz, L. H., Binegar, D., Weiner, M. F., Lipton, A., & Cullum, C. M. (2006). Patterns of verbal memory performance in mild cognitive impairment, Alzheimer disease, and normal aging. *Cognitive & Behavioral Neurology*, 19, 79-84.
- Hort, J., O’Brien, J. T., Gainotti, G., Pirttila, T., Popescu, B. O., Rektorova, I., ...Scheltens, P. (2010). EFNS Scientist Panel on Dementia. EFNS guidelines for the diagnosis and management of Alzheimer’s disease. *European Journal of Neurology*, 17, 1236–1248.
- Knopman, D. S., DeKosky, S. T., Cummings, J. L., Chui, H., Corey-Bloom, J., Relkin, N., ...Stevens, J. C. (2001). Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, 56, 1143–1153.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test. Manual*. Leiden: PITS bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual*

- Association Test. Manual.* Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry, 73*, 126-133.
- Meyer, S. R. A., de Jonghe, J. F. M., Schmand, B., & Ponds, R. W. H. M. (2017). The Visual Association Test-Extended: A cross-sectional study of the performance validity measures. *The Clinical Neuropsychologist, 31*, 789-813.
- Meyer, S. R. A., Spaan, P. E. J., Boelaarts, L., Ponds, R. W. H. M., Schmand, B., & de Jonghe, J. F. M. (2016). Visual associations cued recall A paradigm for measuring episodic memory decline in Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 23*, 566-577.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B.T., Jack, C. R. Jr., Kawas, C. H., ...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, 7*, 263-269.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment. *Archives of Neurology, 56*, 303-308.
- Rey, A. (1958). *L'examen Clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Saan, R. J., & Deelman, B. G. (1986). *15-Woorden Test A en B* [15-Word Test A and B]. In Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (Eds.) (2012). *Handboek neuropsychologische diagnostiek* [Handbook neuropsychological assessment] (2nd ed., p.p. 267-282). Amsterdam: Pearson Assessment and Information B.V.
- Schmand, B., Deelman, B. G., Hooijer, Ch., Jonker, C., & Lindeboom, J. (1996). De item-

- reeks van de cognitieve screening test vergeleken met die van de mini-mental state examination [The items of the cognitive screening test compared to those of the mini-mental state examination]. *Tijdschrift voor Gerontologie en Geriatrie*, 27, 29-33.
- Sorbi, S., Hort, J., Erkinjuntti, T., Fladby, T., Gainotti, G., Gurvit, H., ...Scheltens, P. (2012). EFNS Scientist Panel on Dementia and Cognitive Neurology. EFNS-ENS Guidelines on the diagnosis and management of disorders associated with dementia. *European Journal of Neurology*, 19, 1159–1179.
- Spaan, P. E., Raaijmakers, J. G., & Jonker, C. (2003). Alzheimer's disease versus normal ageing: a review of the efficiency of clinical and experimental memory measures. *Journal of Clinical and Experimental Neuropsychology*, 25, 216–233.
- Van der Elst, W., van Boxtel, M. P., van Breukelen, G. J., & Jolles, J. (2005). Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. *Journal of the International Neuropsychological Society*, 11, 290–302.
- Verhaeghen, P. & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, 122, 231–249.





PART III
PERFORMANCE VALIDITY OF
EPISODIC MEMORY

6

Chapter 6

THE VISUAL ASSOCIATION TEST- EXTENDED: A CROSS-SECTIONAL STUDY OF THE PERFORMANCE VALIDITY MEASURES

Sascha R.A. Meyer, Jos F.M. de Jonghe, Ben Schmand, and Rudolf W.H.M. Ponds

The Clinical Neuropsychologist, 2017; 31: 789-813

Abstract

Given the hazards of knowledge about performance validity tests (PVTs) being proliferated amongst the general public, there is a continuous need to develop new PVTs. The purpose of these studies was to validate the newly developed Visual Association Test - Extended (VAT-E). The VAT-E consists of 24 pairs of line drawings; it is partly based on Green's Word Memory Test (WMT) paradigm. In study 1, we compared VAT-E total scores of healthy controls ($n = 226$), patients with mild cognitive impairment (MCI) ($n = 76$), patients with Alzheimer's disease (AD) ($n = 26$), and persons instructed to feign memory deficit ($n = 29$). In study 2, we compared litigating patients classified by Slick's criteria as Malingering of Neurocognitive Dysfunction (MND) ($n = 26$) or non-MND ($n = 67$). In addition, we compared the VAT-E to the Test of Memory Malingering (TOMM) (study 1) and the WMT (study 2). Results showed that the VAT-E differentiated patients with MCI (specificity 93% - 100%) or patients with AD (specificity 92% - 100%) from persons instructed to feign (sensitivity 86% - 100%). The VAT-E also differentiated MND from non-MND (sensitivity 54%, specificity 97%). The VAT-E was in perfect agreement with the TOMM in classifying healthy controls and persons instructed to feign, and it was in moderate agreement with the WMT in classifying non-MND and MND. Preliminary evidence shows that the VAT-E may be a useful PVT based on the ability to differentiate between those with genuine memory impairment, persons instructed to feign memory impairment, and a group suspected of malingering cognitive deficits.

Introduction

The results of neuropsychological tests may be distorted by patients who exaggerate or fabricate cognitive deficits. Prevalence of such non-credible performance is estimated between 22% and 40% for patients involved in litigation (Larrabee, 2007; Mittenberg, Patton, Canyock, & Condit, 2002). Cognitive test scores that do not accurately reflect a patient's true capabilities may lead to an incorrect diagnosis (Bush et al., 2005; Guilmette, 2013; Heilbronner, et al., 2009). Misdiagnosis can be reduced if performance validity tests (PVTs) are used to detect non-credible performance. However, PVTs are vulnerable to coaching (Suhr & Gunstad, 2007), and people can find essential information on PVTs on the Internet (Bauer & McCaffrey, 2006). Given the hazards of knowledge about PVT paradigms being proliferated amongst the general public, there is a continuous need for new PVTs.

Therefore, we constructed the Visual Association Test - Extended (VAT-E). It is partly based on the well-validated and widely used Word Memory Test paradigm (WMT; Green, 2003). However, the VAT-E is visual-based, as it presents pictures instead of words. Frequently used visual-based PVTs are the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Non-Verbal Medical Symptom Validity Test (NV-MSVT; Green, 2008; Schroeder, Martin, & Odland, 2016; Young, Roper, & Arentsen, 2016). Sensitivity for non-credible responding, as well as specificity for dementia of the NV-MSVT is higher than the TOMM (Green, 2011; Armistead-Jehle & Gervais, 2011; Henry, Merten, Wolf, & Harth, 2010; Merten, Bossink, & Schmand, 2007). During the performance validity measures immediate recognition (IR) and delayed recognition (DR), the TOMM, as well as the NV-MSVT uses side by side presentation of a target picture and a foil picture. However, the VAT-E presents pictures of objects and animals that are linked through an interaction, e.g., a dog riding a bicycle, during IR and DR (See Figure 1). Furthermore, the TOMM does not employ additional measures of memory, as opposed to the NV-MSVT and the VAT-E, which employ

paired associate recall (PA) and free recall (FR), and PA, FR, and multiple choice (MC) respectively. In addition, the VAT-E employs the improvement from FR to MC as an embedded performance validity measure.

We report on two studies that were aimed at validating the VAT-E. In the first study, we hypothesized that the VAT-E discriminates between persons instructed to feign memory deficit on the one hand, and healthy controls, patients with mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD) on the other hand. In the second study, we hypothesized that the VAT-E discriminates between litigating patients classified as Malingering of Neurocognitive Dysfunction (MND; Slick, Sherman, & Iverson, 1999) and litigating patients classified as non-MND.

STUDY 1

Methods

Participants

We included 226 healthy controls, 76 patients with MCI, 26 patients with mild AD, and 29 persons instructed to feign memory deficit. The MCI group contained 61 patients with amnesic MCI and 15 patients with multiple domain MCI. All participants had intact or corrected to normal visual and auditory acuity. The MCI and AD patients were recruited from the outpatient memory clinic of the department of geriatric medicine of the Northwest Medical Center in Alkmaar, the Netherlands. The healthy controls and persons instructed to feign were relatives of the geriatric outpatients or were recruited from the informal network of the researchers. All these participants gave written informed consent. Healthy controls and persons instructed to feign were eligible for inclusion if they scored 13 or higher out of 20 on the Cognitive Screening Test (CST; De Graaf & Deelman, 1991). Included MCI and AD patients had a comprehensive geriatric evaluation, neuropsychological examination and

neuroimaging. All relevant diagnostic information was discussed in a multidisciplinary meeting of five geriatricians, a neurologist, a clinical neuropsychologist, and an old-age psychiatrist. Diagnosis of AD or MCI was based on internationally accepted criteria (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision [DSM-IV-TR]; American Psychiatric Association, 2000; Dubois et al., 2007; Petersen et al., 2001). The global Clinical Dementia Rating (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982; Morris, 1993) score was 0.5 for all MCI patients and 1 for all AD patients. Hippocampal atrophy was grade 2 or more for 46 MCI patients and 20 AD patients (Scheltens et al., 1992). Mean total score on delayed recall of the Rey Auditory Verbal Learning Test (RAVTL; Rey, 1964; Schmidt, 1996) was 1.95 ($SD = 1.8$, Range = 0-6) for MCI patients, and 0.91 ($SD = 1.5$, Range = 0-6) for AD patients. Exclusion criteria for all participants were severe traumatic brain damage and brain disease (such as brain tumor, epilepsy, multiple sclerosis, Parkinson's disease), severe psychiatric disorder, delirium or history of alcohol or drug abuse. The study was approved by the medical ethical committee of the province North-Holland, the Netherlands.

Visual Association Test-Extended

The VAT-E is based on items of the Visual Association Test (VAT; Lindeboom & Schmand, 2003; Lindeboom, Schmand, Meyer, & de Jonghe, 2014). The VAT is an episodic memory test that presents pictures of a target stimulus and a cue interacting with one another, such as a dog (target) riding a bicycle (cue). Subsequently, the target has to be remembered by way of PA or MC. The VAT-E was constructed by joining VAT 12-item version AB with CD. For the CD version, a four-alternative recognition trial is available, which we used for the MC trial of the VAT-E.

Figure 1 shows samples of each trial of the VAT-E. First, in two learning trials, the subject is instructed to look at and remember 24 pictures, i.e., black and white line drawings of common objects and animals, e.g., a bicycle. These pictures are presented separately at a pace of four seconds per picture. Next, in the IR trial, each picture is shown again with an associated picture, linking these through an interaction, e.g., a dog riding the bicycle. The subject is instructed to name both pictures aloud, inducing incidental learning of the association of the pictures, and then to point out the picture that had to be remembered during the learning trials, i.e., the bicycle. After a 15-minute interval, the DR trial is administered in the same order, with the same instruction. During the 15-minute interval of the VAT-E, non-visual or non-memory tests can be administered. The number of responses in agreement between IR and DR is calculated for the consistency measure (CNS). The total score on IR, DR and CNS ranges from 0 to 24. Next, in the PA trial, the 24 pictures from the learning trials are now presented one by one as cues, e.g., the bicycle, and the subject is instructed to recall the associated picture, e.g., the dog. The total score on PA ranges from 0 to 24. Next, in the FR trial, the subject is instructed to recall as many pictures as possible from all the pictures that were shown, i.e., pictures from the learning trials and associated pictures. The total score on FR ranges from 0 to 48. Finally, in the MC trial, the last 12 pictures from the learning trials are again presented one by one as cues, but now the subject is instructed to point out the associated picture choosing from four alternatives. The total score on MC ranges from 0 to 12. During all the trials, the test administrator gives feedback about the subject's performance on each item, i.e., "this is correct" or "this is wrong". In case of a wrong answer, the test administrator always presents the correct response, except during FR, to enable additional learning.

The VAT-E employs four performance validity measures. The first three are IR, DR and CNS. These capitalize on the well-researched finding that healthy controls and patients

with genuine memory impairment show ceiling effects on multiple choice tests of recognition, which are not shown by persons exaggerating or fabricating memory impairment (Green, Iverson, & Allen, 1999; Tombaugh, 1996). Measures of memory are subsequent PA, FR, and MC.

The fourth VAT-E performance validity measure is embedded, as it compares the performance on the measure of memory FR to the performance on the measure of memory MC (FR-MC) in relation to known patterns of memory functioning (Green, 2003; Rogers, 2008; Slick et al., 1999). The FR-MC performance validity measure capitalizes on the well-researched finding that the retrieval of test material improves for AD or MCI patients from recall to recognition (Bäckman, Jones, Berger, Laukka, & Small, 2005). Persons exaggerating or fabricating memory impairment are unaware of this memory function pattern, and their FR versus MC scores show much less improvement than the scores of AD or MCI patients (Green, 2003; Haines & Norris, 1995; Rogers, 2008; Slick et al., 1999). Therefore, if a person performs above the mean FR total score, but below the mean MC total score of MCI patients, this is not in line with known patterns of memory functioning, and thus, suspect of non-credible performance.

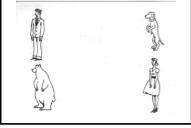
| Trial | Sample of trial |
|-----------------------------|---|
| 6. Learning |  |
| 7. Learning |  |
| 8. Immediate recognition |  |
| 9. Delayed recognition |  |
| 10. Paired associate recall |  |
| 6. Free recall | - |
| 8. Multiple choice |   |

Figure 1. Sample of VAT-E trials.

Note. VAT-E = Visual Association Test-Extended.

Additional Measures

The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a visual-based PVT. It presents two learning trials, each followed by an IR trial. After a 15-minute interval, a DR trial is administered. A failure on the TOMM is defined by a score below the published cut-off score on one of the performance validity measures.

The Cognitive Screening Test (CST; De Graaf & Deelman, 1991) is used to screen for global cognitive impairment. It includes 20 questions concerning orientation in time, place

and person, and elementary factual knowledge. The total score ranges from 0 to 20, with impairment defined by a score of less than 13. The CST is comparable to the Mini-Mental State Examination (MMSE; Folstein, Folstein, & HcHugh, 1975) in predicting dementia (Schmand, Deelman, Hooijer, Jonker, & Lindeboom, 1996).

The Clinical Dementia Rating (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982; Morris, 1993) is a clinical global scale to indicate the presence and stage the severity of AD and other dementias. The severity of dementia is rated on five levels of global cognitive impairment, i.e., none (0), questionable (0.5), mild (1), moderate (2), and severe (3).

The Dutch version of the RAVLT was used (15WT; Saan & Deelman, 1986). It presents a list of 15 words that are read out loud to the subject in five repeated trials, which have to be recalled immediately after each presentation and after a 20-minute time interval, i.e., delayed recall.

Education was scored in accordance with the Dutch scoring system Verhage (Verhage, 1964), which by and large overlaps with the International Standard Classification of Education 1997 scales (UNESCO, 2006). The scores range from one to seven, with one corresponding to unfinished elementary school and seven to university degree education.

Design and Procedure

We compared VAT-E total scores on the four performance validity measures of patients with genuine memory impairment to persons instructed to feign memory deficit, and compared classification according to optimal cut-off scores of the VAT-E to the TOMM. In our simulation design (Rogers, 2008), we randomly assigned healthy participants to either the healthy control or feigning group. The CST was administered before the feigning instruction. Participants in the feigning group were given the instruction to feign subtle memory problems, in order to demonstrate their non-suitability to return to work after an accident. To

maximize motivation, participants were told that they would be rewarded with € 10,- if they did this in a way the researchers considered credible (in reality, all subjects received this reward). Subsequently, the VAT-E and TOMM were administered. Participants in the healthy control group were asked to perform these tests to the best of their abilities. Patients with MCI or AD underwent the VAT-E and CST as part of the standardized neuropsychological assessment.

Data Analysis

Demographic characteristics. We investigated the effect of gender, age, and education on the four VAT-E performance validity measures by use of Pearson's chi-square test, Kruskal-Wallis tests, Mann-Whitney tests, t-tests and Spearman's and Pearson's correlations, as appropriate.

Predictive validity. We compared VAT-E total scores of MCI or AD patients to persons instructed to feign memory deficits in a receiver operating characteristic (ROC) analysis for a range of IR, DR and CNS cut-off scores. We investigated if persons instructed to feign memory deficit improved less than MCI patients, by comparing FR and MC mean total scores between these groups using Mann-Whitney tests and Cohen's *d*. We set the FR-MC cut-off score at the upper 95% confidence interval of the FR mean and the lower 95% confidence interval of the MC mean of MCI patients. We used binomial distribution (Agresti & Franklin, 2007) to calculate below chance level (Frederick & Speed, 2007; Grote & Hook, 2007; Heilbronner, et al., 2009; Slick, et al., 1999) performance on MC. In the present case of 12 four-choice items, a score of 0 is below chance ($p = 0.03$, one-tailed). We calculated sensitivity and specificity for a range of FR-MC cut-off scores of the groups. We selected optimal cut-off scores for all four VAT-E performance validity measures, using the maximum value of Youden's *J* statistic (sensitivity + specificity – 1; Youden, 1950), which is a

summary measure of the ROC curve and corresponds to the optimal balance between sensitivity and specificity. The optimal cut-off score is determined by calculating J for all cut-off scores of a measure, and subsequently selecting the maximum. In our analyses sensitivity was defined by persons instructed to feign memory deficit, and specificity was defined by the MCI or AD group.

Concurrent validity. We used the optimal cut-off scores based on the MCI group, to investigate whether VAT-E was in agreement with TOMM in classifying health controls and persons instructed to feign memory deficit, using Cohen's kappa.

Results

Table 1 shows the demographic characteristics of the groups. The groups differed in age ($H(3) = 145.4, p < .01$), and education ($H(3) = 71.2, p < .01$). Table 2 shows test performances of the groups. Scores on the VAT-E performance validity measures IR, DR, and CNS showed a ceiling effect for healthy controls, MCI patients, and AD patients. For HCs, age ($r(226) = -.39, p < .01$) and education ($r(226) = .34, p < .01$) had an effect on FR, and education correlated weakly with MC ($r(226) = .20, p < .01$).

Table 1. Demographic characteristics according to group (both studies)

| | Healthy controls | MCI | AD | Instructed feigning (<i>n</i> = 29) | non-MND | MND |
|------------|--------------------------------|--------------------------|--------------------------|---|--------------------------|--------------------------|
| Female (%) | (<i>n</i> = 226) 62.4 | (<i>n</i> = 76) 48.7 | (<i>n</i> = 26) 53.8 | 58.6 | (<i>n</i> = 67) 26.9 | (<i>n</i> = 26) 32.8 |
| Age | 53.7 (16.4) | 76.2 (6.0) | 79.5 (6.7) | 60.0 (7.8) | 49.6 (10.1) | 47.2 (10.1) |
| Education | 5.4 (1.2) | 4.3 (1.1) | 4.3 (1.1) | 5.7 (0.8) | 5.4 (1.3) | 4.2 (1.6) |
| CST | 19.1 ^a (1.2; 15-20) | 15.9 (2.1; 8-20) | 13.6 (2.7; 6.5-18) | 19.2 (1.1; 17-20) | - | - |

Note. Scores were rounded up if the next smallest digit was greater than or equal to five; MCI = amnesic and multi domain mild cognitive impairment; AD = Alzheimer's Disease; MND = Malingering of Neurocognitive Dysfunction; number of subjects (i.e., *n*); female (proportion of females); age in years (mean and standard deviation); education in accordance with the Verhage system (mean and standard deviation); CST = Cognitive Screening Test (mean, standard deviation and range of scores).

^a *n* = 158.

Table 2. Test performances according to group (both studies)

| | Healthy controls (<i>n</i> = 226) | MCI (<i>n</i> = 76) | AD (<i>n</i> = 26) | Instructed feigning (<i>n</i> = 29) | non-MND (<i>n</i> = 67) | MND (<i>n</i> = 26) |
|-------------|---------------------------------------|-------------------------|------------------------|---|--------------------------------|---------------------------------|
| VAT-E – IR | 24.0 (0.2; 22-24) | 23.6 (0.7; 22-24) | 23.3 (1.1; 21-24) | 16.5 (2.5; 12-21) | 24.0 (0.2; 23-24) | 19.3 (6.2; 5-24) |
| VAT-E – DR | 24.0 (0.2; 22-24) | 23.4 (1.3; 16-24) | 22.3 (1.7; 20-24) | 15.3 (2.7; 11-20) | 23.9 (0.7; 19-24) | 18.8 (6.4; 5-24) |
| VAT-E – CNS | 24.0 (0.3; 22-24) | 23.3 (1.3; 16-24) | 22.3 (1.7; 19-24) | 16.1 (3.4; 9-21) | 23.9 (0.6; 20-24) | 21.0 (3.8; 12-24) |
| VAT-E – PA | 18.3 (5.1; 1-24) | 7.8 (6.5; 0-23) | 4.3 (4.3; 0-18) | 10.7 (3.3; 3-19) | 18.3 (6.1; 0-24) | 13.1 (7.0; 0-24) |
| VAT-E – FR | 26.7 (8.5; 5-48) | 6.1 (4.0; 0-17) | 3.6 (4.2; 0-16) | 15.5 (5.6; 4-28) | 20.6 (7.4; 6-42) | 10.3 (5.2; 2-21) |
| VAT-E – MC | 11.9 (0.3; 10-12) | 9.6 (2.7; 2-12) | 7.2 (3.5; 1-12) | 6.5 (1.7; 4-10) | 11.7 (0.8; 8-12) | 9.0 (4.1; 0-12) |
| TOMM – T2 | 49.9 ^a (0.4; 48-50) | - | - | 33.7 (5.8; 20-45) | - | - |
| TOMM – DR | 50.0 ^a (0.2; 49-50) | - | - | 35.2 (4.8; 24-46) | - | - |
| WMT – IR | - | - | - | - | 92.9 (8.8; 65-100) | 59.8 ^b (22.5; 25-95) |
| WMT – DR | - | - | - | - | 92.2 (10.0; 63-100) | 56.6 ^b (22.7; 15-88) |
| WMT – CNS | - | - | - | - | 89.4 (12.1; 53-100) | 64.8 ^b (12.2; 43-83) |
| RDS | - | - | - | - | 8.7 ^c (1.8; 3-15) | 5.9 ^d (2.4; 2-11) |
| ASTM | - | - | - | - | 85.8 ^e (3.3; 76-90) | 70.3 ^f (21.2; 9-86) |

Note. Scores were rounded up if the next smallest digit was greater than or equal to five; MCI = amnesic and multi domain mild cognitive impairment; AD = Alzheimer’s Disease; MND = Malingering of Neurocognitive Dysfunction; VAT-E = Visual Association Test-Extended (mean, standard deviation and range of scores); TOMM = Test of Memory Malingering (mean, standard deviation and range of scores); WMT = Word Memory Test (mean, standard deviation and range of scores); RDS = Reliable Digit Span; ASTM = Amsterdam Short-Term Memory test; IR = immediate recognition; DR = delayed recognition; CNS = consistency measure; PA = paired associate recall; FR = free recall; MC = multiple choice; T2 = Trial 2.

^a *n* = 45; ^b *n* = 21; ^c *n* = 23; ^d *n* = 66; ^e *n* = 35; ^f *n* = 11.

Figure 2 shows the score profile of VAT-E scores in percentage correct for each group. Table 2 shows that the mean FR score of persons instructed to feign was higher than that of MCI patients, with a large effect size ($U = 195.5, p < .001, \text{Cohen's } d = 1.9$). As predicted, the mean MC score of persons instructed to feign was lower than that of MCI patients, again with a large effect size ($U = 371.5, p < .001, \text{Cohen's } d = 1.4$). Table 3 shows the diagnostic accuracy for a range of IR, DR, CNS and FR-MC cut-off scores of the groups. Based on the MCI group the optimal cut-off scores were set at $IR \leq 21, DR \leq 20, CNS \leq 21,$ and $FR \geq 7$ in combination with $MC \leq 9$. Based on the AD group the optimal cut-off scores were set at $IR \leq 20, DR \leq 19, CNS \leq 19,$ and $FR \geq 7$ in combination with $MC \leq 9$. There was perfect agreement between VAT-E and TOMM ($\text{kappa} = 1$).

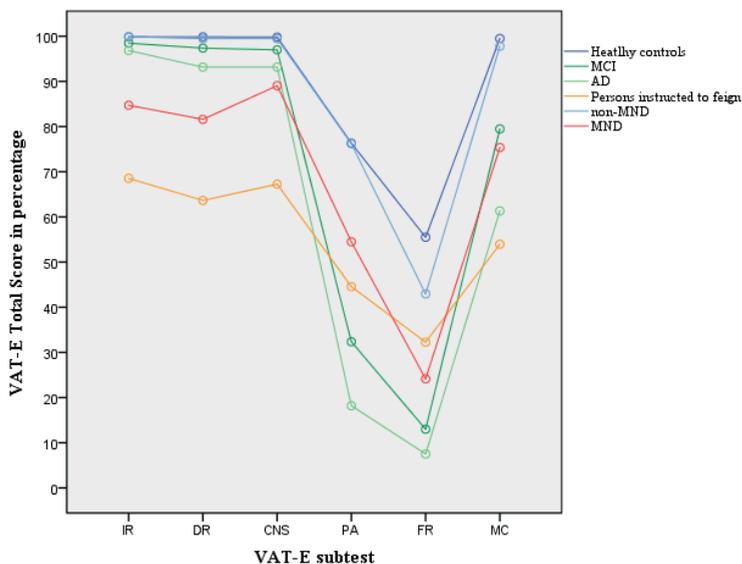


Figure 2. Score profiles of VAT-E total scores as percentage correct scores (both studies) for healthy controls ($n = 226$), MCI patients ($n = 76$), AD patients ($n = 26$) persons instructed to feign memory deficit ($n = 29$), non-MND ($n = 67$), and MND ($n = 26$).

Note. VAT-E = Visual Association Test-Extended; MCI = amnesic and multi domain mild cognitive impairment; AD = Alzheimer’s Disease; MND = Malingering of Neurocognitive Dysfunction; number of subjects (i.e., n); IR = immediate recognition; DR = delayed recognition; CNS = consistency measure; PA = paired associate recall; FR = free recall; MC = multiple choice.

Table 3. Diagnostic accuracy (study 1) of VAT-E performance validity measures for MCI patients ($n = 76$), AD patients ($n = 26$), and persons instructed to feign memory deficit ($n = 29$).

| VAT-E | AUC for MCI vs. Instructed feigning (95% CI) | AUC for AD vs. Instructed feigning (95% CI) | Cut-off score | Sensitivity for Instructed feigning | Specificity for MCI | Specificity for AD |
|-------|---|--|------------------|--|---------------------------|--------------------------|
| IR | 1 | .99 (.98-1) | ≤ 23 | 100 % | 71 % | 62 % |
| | | | ≤ 22 | 100 % | 92 % | 81 % |
| | | | ≤ 21 | 100 % | 100 % | 85 % |
| | | | ≤ 20 | 93 % | 100 % | 100 % |
| | | | | | | |
| DR | .99 (.99-1) | .98 (.96-1) | ≤ 23 | 100 % | 68 % | 39 % |
| | | | ≤ 22 | 100 % | 83 % | 50 % |
| | | | ≤ 21 | 100 % | 92 % | 65 % |
| | | | ≤ 20 | 100 % | 99 % | 73 % |
| | | | ≤ 19 | 90 % | 99 % | 100 % |
| CNS | .99 (.98-1) | .97 (.93 -1) | ≤ 23 | 100 % | 58 % | 39 % |
| | | | ≤ 22 | 100 % | 83 % | 50 % |
| | | | ≤ 21 | 100 % | 96 % | 62 % |
| | | | ≤ 20 | 93 % | 96 % | 81 % |
| | | | ≤ 19 | 86 % | 99 % | 96 % |
| FR-MC | - | - | ≥ 5 - ≤ 11 | 97 % | 65 % | 73 % |
| | | | ≥ 6 - ≤ 10 | 97 % | 82 % | 85 % |
| | | | ≥ 7 - ≤ 9 | 90 % | 93 % | 92 % |
| | | | ≥ 8 - ≤ 8 | 79 % | 99 % | 92 % |
| | | | ≥ 9 - ≤ 7 | 55 % | 99 % | 100 % |

Note. MCI = amnesic and multi domain mild cognitive impairment; AD = Alzheimer’s Disease; number of subjects (i.e., n); VAT = Visual Association Test-Extended; AUC = area under the receiver operating curve; CI = confidence interval; IR = immediate recognition; DR = delayed recognition; CNS = consistency measure; FR-MC = free recall in combination with multiple choice.

STUDY 2

Methods

Participants

A convenience sample of litigating patients (archival data) was classified as non-MND or MND (Slick et al., 1999). We included 67 litigating patients classified as non-MND, and 26 litigating patients classified as MND. All participants had intact or corrected to normal visual and auditory acuity. Table 4 shows the description of the litigation group. The VAT-E was administered to the most recent patients, from the moment we constructed the test. All litigating patients were referred for neuropsychological examination by insurance physicians or company doctors. Exclusion criterion for all participants was severe brain damage. Two patients who had undergone neurosurgery and one patient with dyslexia, who was administered the WMT, were excluded.

Table 4. Description of the litigation group ($n = 93$) (study 2).

| Patients | <i>n</i> |
|--------------------------------|----------|
| Cerebrovascular accident | 16 |
| Depression | 13 |
| Mild traumatic brain injury | 10 |
| Post-traumatic stress disorder | 10 |
| Job burnout | 6 |
| Whiplash | 6 |
| Anxiety | 5 |
| Chronic pain | 5 |
| Depression and anxiety | 4 |
| Chronic fatigue | 3 |
| Bipolar disorder | 2 |
| Obstructive sleep apnea | 2 |
| Personality disorder | 2 |
| Neurotoxicity | 2 |
| Fibromyalgia | 1 |
| Multiple sclerosis | 1 |
| Epilepsy | 1 |
| Atrial fibromyalgia | 1 |
| Autism spectrum disorder | 1 |
| Pulmonary embolism | 1 |
| Migraine | 1 |

Note. Number of subjects (i.e., *n*).

Measurements

Some measurements were the same as in study 1, i.e., VAT-E, TOMM, and education as scored in accordance with the Dutch scoring system Verhage.

The computerized Word Memory Test (WMT; Green, 2003) is a verbal-based PVT. It presents two learning trials, followed by administration of IR. After a 30-minute interval, DR is administered, and the consistency between these trials is measured as CNS. After these trials MC, PA and FR are administered. The long delayed free recall condition of the test was not used.

The Amsterdam Short-Term Memory test (ASTM; Schmand & Lindeboom, 2004) presents a list of five words, that has to be read aloud by the subject. Next, the subject has to add or subtract numbers. Then, the subject has to recognize three words from the previous list out of a list of five words. The test consists of 30 items.

The Non-Verbal Medical Symptom Validity Test (NV-MSVT; Green, 2008) presents two learning trials of 10 artist-drawn colored image pairs, after which a series of two forced-choice recognition memory trials are administered. These are followed by PA and FR trials.

The WAIS-III Digit Span test (Wechsler, 1997) was administered to establish the Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994), which is calculated by summing the longest string of digits forward and backward passed over two trials. A failure on the RDS is defined by a score at or below the published cut-off score (Greve et al., 2010).

Scores on the PVTs are classified as non-credible, if the score on one or more of the tests performance validity measures is below the published cut-off score.

Design and Procedure

We compared VAT-E total scores on the four performance validity measures of litigating patients classified as non-MND to litigating patients classified as MND, and compared classification according to optimal cut-off scores of the VAT-E to MND classification and the WMT. Litigating patients were classified as MND (Slick et al., 1999) if they had a substantial external incentive and scored either below chance on one memory PVT or below published

cut-off scores on two or more (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2008; Larrabee, Greiffenstein, Greve, & Bianchini, 2007; Meyers et al., 2014; Vickery et al., 2004) of the following memory PVTs: ASTM, NV-MSVT, RDS, TOMM, and WMT. The WMT was administered to all non-MND and 21 MND. The RDS was administered to 66 non-MND and 23 MND. The ASTM was administered to 35 non-MND and 11 MND. The NV-MSVT was administered to 3 MND. The TOMM was administered to 2 MND. The average number of PVTs administered to non-MND and MND was 2.5 and 2.3 respectively. Litigating patients were administered these tests and the VAT-E as part of the standardized neuropsychological assessment, but the VAT-E was not used for MND classification. The PVTs were administered in a fixed order, i.e., WMT, ASTM, VAT-E, and RDS. The WMT or ASTM were substituted by the NV-MSVT or TOMM respectively, in case of the necessity for language independent assessment.

Data Analysis

Predictive validity. We compared VAT-E total scores of litigating patients classified as non-MND to litigating patients classified as MND in a ROC analysis for a range of IR, DR, and CNS cut-off scores. Litigating patients scoring below chance level on MC, were a priori included as showing non-credible performance on FR-MC. We calculated sensitivity and specificity for a range of FR-MC cut-off scores. We investigated whether the optimal cut-off scores of the VAT-E based in the MCI group, differentiated MND from non-MND by calculating sensitivity, specificity, positive predictive value, and negative predictive value.

Concurrent validity. We used the optimal cut-off scores based on the MCI group, to investigate whether VAT-E was in agreement with WMT in classifying non-MND and MND, using Cohen's kappa.

Results

Table 1 shows the demographic characteristics of the groups. The groups differed in education ($U = 498, p < .01$). Figure 2 shows the score profile of VAT-E scores in percentage correct for each group. Table 2 shows test performances of the groups. Scores on the VAT-E performance validity measures IR, DR, and CNS showed a ceiling effect for non-MND. Two MND patients scored below chance level on MC. Table 5 shows the diagnostic accuracy for a range of IR, DR, CNS and FR-MC cut-off scores of the groups.

Table 5. Diagnostic accuracy (study 2) of VAT-E performance validity measures for litigating patients classified as non-MND ($n = 67$) or MND ($n = 26$).

| VAT-E | AUC for non-MND vs. MND (95% CI) | Cut-off score | Sensitivity for MND | Specificity for non-MND |
|-------|---|------------------|------------------------|----------------------------|
| IR | .78 (.66-.91) | ≤ 23 | 58 % | 97 % |
| | | ≤ 22 | 54 % | 100 % |
| | | ≤ 21 | 46 % | 100 % |
| | | ≤ 20 | 39 % | 100 % |
| DR | .79 (.67-.91) | ≤ 23 | 62 % | 93 % |
| | | ≤ 22 | 54 % | 99 % |
| | | ≤ 21 | 50 % | 99 % |
| | | ≤ 20 | 42 % | 99 % |
| | | ≤ 19 | 42 % | 99 % |
| CNS | .78 (.66-.91) | ≤ 23 | 62 % | 91 % |
| | | ≤ 22 | 54 % | 99 % |
| | | ≤ 21 | 39 % | 99 % |
| | | ≤ 20 | 31 % | 99 % |
| | | ≤ 19 | 23 % | 100 % |
| FR-MC | - | ≥ 5 - ≤ 11 | 42 % | 87 % |
| | | ≥ 6 - ≤ 10 | 27 % | 93 % |
| | | ≥ 7 - ≤ 9 | 19 % | 97 % |
| | | ≥ 8 - ≤ 8 | 15 % | 100 % |
| | | ≥ 9 - ≤ 7 | 8 % | 100 % |

Note. MND = Malingering of Neurocognitive Dysfunction; number of subjects (i.e., n); VAT-E = Visual Association Test-Extended; AUC = area under the receiver operating curve; CI = confidence interval; IR = immediate recognition; DR = delayed recognition; CNS = consistency measure; FR-MC = free recall in combination with multiple choice.

The VAT-E differentiated MND from non-MND with sensitivity of 54%, specificity of 97%, positive predictive value of 88%, and negative predictive value of 84%. There was moderate agreement between VAT-E and WMT ($\kappa = 0.36$ (95% CI, .19 to .53), $p < .01$), with more than twice as many litigating patients failing the WMT, twelve of whom were MND and twelve were non-MND.

Discussion

These two studies examined the predictive and concurrent validity of the VAT-E, a newly developed performance validity test. Four performance validity measures of the VAT-E proved highly effective in differentiating MCI or AD patients from persons instructed to feign, and the test showed adequate sensitivity and specificity for MND in liability claimants. Thus, the VAT-E shows promise as a valid performance validity measure.

Strong points of our studies are the use of a simulation design, i.e., comparison of persons instructed to feign with bona fide memory patients, as well as a known-groups comparison, i.e., MND versus non-MND, and the comparison of the VAT-E with the WMT and the TOMM. The simulation design enabled us to set optimal cut-off scores with high control over internal validity. Furthermore, we addressed external validity in our known-groups comparison by testing these cut-off scores in a real-life sample of litigating patients classified according to the Slick-criteria, which are widely used in research (Greve & Bianchini, 2006; Greve, Bianchini, & Doane, 2006). Since we also compared the VAT-E classifications with two well-validated and widely used PVTs, we consider our findings to be robust.

Our findings are in line with results from previous studies, which show that sensitivity and specificity of memory PVTs vary depending on the use of visual or verbal stimuli. In general, sensitivity is somewhat lower in memory PVTs based on pictures, such as the

TOMM and the NV-MSVT, than in memory PVTs based on words, such as the WMT, but specificity is higher (Gervais, Rohling, Green, & Ford, 2004; Green, 2007; Green, 2008; Green, Flaro, Brockhaus, & Montijo, 2012; Greve, Ord, Curtis, Bianchini, & Brennan, 2008). A possible explanation for this finding is that pictures are easier to recognize than words (Shepard, 1967; Paivio, Rogers, & Smythe, 1968), since pictures are more often coded in a dual process than words, i.e., visually as well as verbally (Paivio, 1969; Paivio et al., 1968). This facilitates recognition of pictures more than recognition of words, thus, on the one hand specificity for genuinely memory impaired patients will be high, but on the other hand, informed malingerers might interpret a test that is too easy as a PVT. Furthermore, previous research shows that sensitivity of performance validity measures administered earlier in a neuropsychological assessment is higher (Guilmette, Whelihan, Hart, Sparadeo, & Buongiorno, 1996). Because the VAT-E was administered last of the stand-alone PVTs, i.e., ASTM, NV-MSVT, TOMM or WMT, this could have had an effect on its sensitivity for MND. However, if this had an effect, than sensitivity will be higher in future earlier administrations. Thus, the sensitivity of the VAT-E possibly is an underestimation of its potential predictive power.

A limitation that needs to be discussed is the generalizability of the optimal VAT-E cut-off scores established in the simulation design. It could be argued that the behavior elicited by the simulation instruction does not mimic behavior in the real world, because persons instructed to feign lacked the external incentive. The € 10,- reward may have been insufficient. However, the optimal cut-off scores proved adequate in differentiating MND from non-MND in litigating patients, and hence we consider them to be adequate for clinical use.

Another limitation is whether it is certain that MND actually constitutes exaggeration or fabrication of memory impairment. Unfortunately, the gold standard would be patients

admitting MND, which they will rarely do. Instead, we used psychometric indicators, which are currently the accepted approach for identifying MND (Bush et al., 2005; Guilmette, 2013; Heilbronner, et al., 2009). Also, we applied these indicators conservatively by only classifying litigating patients as MND if they scored below chance on one memory PVT (Slick et al., 1999) or below cut-off scores on two or more PVTs (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2008; Larrabee et al., 2007; Meyers et al., 2014; Vickery et al., 2004) and had a substantial external incentive (Slick et al., 1999). Thus, we consider it highly probable that our MND classification indeed reflected exaggeration or fabrication of memory impairment.

The newly developed VAT-E enables valid interpretation of test findings. Currently available PVTs have been in use for more than 20 years. Over time, patients who wanted to exaggerate or fabricate memory impairment could become knowledgeable about PVTs or about how they work. Unfortunately, such information is readily accessible on the internet (Bauer & McCaffrey, 2006) and some lawyers involved in litigation coach their clients on PVTs (Suhr & Gunstad, 2007). As PVTs are vulnerable to coaching (Suhr & Gunstad, 2007) and prevalence of non-credible performance is high amongst litigating patients (Larrabee, 2007; Mittenberg et al., 2002), this potentially leads to the misinterpretation of test findings and, consequently, to an incorrect diagnosis. However, since the VAT-E is a new PVT, it is less vulnerable to coaching.

We report a range of cut-off scores of the groups to allow the clinician to determine the relative cost or benefit of a given cut-off score in a particular setting. However, we consider the optimal cut-off scores determined in our research are adequate for clinical use. First, the optimal cut-off scores based on the MCI group adequately differentiate MND from non-MND in a group of litigating claimants. Second, the optimal cut-off scores, as determined by the maximum value of Youden's J statistic, correspond to the optimal balance between

sensitivity and specificity. Third, for screening purposes an instrument may allow a relatively high false-positive rate, but instruments with intended higher diagnostic accuracy, such as PVTs, should minimize false-positive classifications. The optimal cut-off scores enable this, because specificity is high. The optimal cut-off scores based on the MCI group could be used for mild memory impairment, such as the frequently claimed mild traumatic brain injury (Caron & Bush, 2013), and the optimal cut-off scores based on the AD group could be used for settings in which moderate to more severe memory impairment is claimed.

Given the hazards of knowledge about PVTs being proliferated amongst the general public, clinical neuropsychologists need new PVTs. Our findings show that the VAT-E is adequate in differentiating non-MND from MND. Furthermore, the stimuli used in the VAT-E are pictures, which enables language-independent administration for the performance validity measures IR, DR and CNS, thus, facilitating testing of a wider range of patients, such as dyslexia patients and non-native speakers. Thus, the VAT-E can be a valuable addition to the toolbox of clinical neuropsychologists who are looking for a new PVT.

Acknowledgments

We would like to thank the Department of Geriatric Medicine of the Northwest Medical Center in Alkmaar, the Netherlands, for the opportunity to conduct our research. Also, we would like to thank Brenda Smolders, MSc, Sylvia Verwer, MSc, Lisa Stel, MSc, Sanne Verhagen, MSc, Merel Oudshoorn, MSc, and Rosalie van der Aa, MSc, for their work on the studies.

Disclosure Statement

S.R.A.M and J.F.M.J are the authors of the Visual Association Test-Extended.

Funding

This work was supported by the Foreest Northwest Medical School [grant number FIO1308].

References

- Agresti, A. & Franklin, C. (2007). *Statistics. The art and science of learning from data.* (p.p. 244-313). New Jersey: Pearson Education Inc.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: Author.
- Armistead-Jehle, P. & Gervais, R. O. (2011). Sensitivity of the Test of Memory Malingering and the Nonverbal Medical Symptom Validity Test: A replication study. *Applied Neuropsychology, 18*, 284-290.
- Bäckman, L., Jones, S., Berger, A., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology, 19*, 520-531.
- Bauer, L. & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the internet: Is test security threatened? *Archives of Clinical Neuropsychology, 21*, 121-126.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ...Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419-426.
- Carone, D. A. & Busch S. S. (Eds.) (2013), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p.p. 1-43). New York: Springer Publishing Company, LLC.
- De Graaf, A. & Deelman, B.G. (1991). *Cognitive Screening Test. Manual for administration and scoring.* Lisse: Swets en Zeitlinger bv.

- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ...Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, *6*, 734-746.
- Frederick, R. I. & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, *14*, 3-11.
- Folstein, M. F., Folstein, S. F., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.
- Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, *19*, 475-487.
- Green, P. (2003). *Green's Word Memory Test for Microsoft Windows: User's manual*. Edmonton, Canada: Green's Publishing Inc.
- Green, P. (2007). Spoiled for choice. Making comparisons between forced-choice effort tests. In Brauer Boone, K. (Ed.) (2007). *Assessment of feigned cognitive impairment* (p.p. 50-77). New York: The Guilford Press
- Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test: User's manual for Microsoft Windows*. Edmonton, Canada: Green's Publishing Inc.
- Green, P. (2011). Comparison between the Test of Memory Malingering (TOMM) and the Nonverbal Medical Symptom Validity Test (NV-MSVT) in adults with disability claims. *Applied Neuropsychology*, *18*, 18-26.
- Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, and NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In Reynolds, C. R. & MacNeill Horton, Jr., A. (Eds.) (2012), *Detection of malingering during head injury litigation* (2nd ed., p.p. 201-219). New

- York: Springer Science+Business Media, LLC.
- Green, P., Iverson, G. L., & Allen, L. (1999) Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*, 813-819.
- Greve, K. W. & Bianchini, K. J. (2006). Classification accuracy of the Portland Digit Recognition Test in traumatic brain injury: Results of a known-groups analysis. *The Clinical Neuropsychologist, 20*, 816-830.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218-224.
- Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-group analysis. *Journal of Clinical and Experimental Neuropsychology, 28*, 1176-1190.
- Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J. E., Curtis, K. L., & Ord, J. S. (2010). The reliable digit span test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist, 24*, 137-152.
- Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22*, 896-918.
- Grote, C. L. & Hook, J. N. (2007). Forced-choice recognition tests of malingering. In Larrabee, G. J. (Ed.) (2007), *Assessment of malingered neuropsychological deficits* (p.p. 44-79). New York: Oxford University Press Inc.
- Guilmette, T. J. (2013). The role of clinical judgement in symptom validity assessment. In Carone, D. A. & Busch S. S. (Eds.) (2013), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p.p. 31-43). New York: Springer Publishing Company, LLC.
- Guilmette, T. J., Whelihan, W. M., Hart, K. J., Sparadeo, F. R., & Buongiorno, G. (1996).

- Order effects in the administration of a forced-choice procedure for detection of malingering in disability claimants' evaluations. *Perceptual and Motor Skills*, 83, 1007-1016.
- Haines, M. E. & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology Review*, 5, 125-148.
- Heilbronner, R. L., Sweet, J. L., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093-1129.
- Henry, M., Merten, T., Wolf, S. A., & Harth, S. (2010). Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients. *Journal of Clinical and Experimental Neuropsychology*, 32, 19-27.
- Hughes, C. P., Berg, L., Danziger, W. L., Coben, L.A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566-572.
- Larrabee, G. J. (Ed.) (2007). *Assessment of malingered neuropsychological deficits* (p.p. 3-13). New York: Oxford University Press Inc.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22, 666-679.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining Diagnostic Criteria for Malingering. In Larrabee, G. J. (Ed.) (2007). *Assessment of malingered neuropsychological deficits* (p.p. 334-371). New York: Oxford University Press Inc.
- Lindeboom, J. & Schmand, B. (2003). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.

- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*, 308-318.
- Meyers, J. E., Miller, R. M., Thompson, L. M., Scalese, A. M., Allred, B. C., Rupp, Z. W., ...Lee, A. J. (2014). Using likelihood ratios to detect invalid performance with performance validity measures. *Archives of Clinical Neuropsychology*, *29*, 224-235.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094-1102.
- Morris, J. C. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, *43*, 2412-2414.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *74*, 241-263.
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*, 137-138.
- Petersen, R. C., Doody, R., Kurz, A. Mohs, R. C., Morris, J.C., Rabins, P. V., ...Winbald, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, *58*, 1985-1992.
- Rey, A. (1964). *L'examen Clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Rogers, R. (Ed.) (2008), *Clinical assessment of malingering and deception* (3rd ed., p.p. 3-13, 14-35, 411-434). New York: The Guilford Press.
- Saan, R. J., & Deelman, B. G. (1986). *15-Woorden Test A en B* [15-Word Test A and B]. In

- Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (Eds.) (2012). *Handboek neuropsychologische diagnostiek* [Handbook neuropsychological assessment] (2nd ed., p.p. 267-282). Amsterdam: Pearson Assessment and Information B.V.
- Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H.C., Vermersch, P., Kuiper, M., ... Kuiper, M. (1992). Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*, 967-972.
- Schmand, B. & Lindeboom, J. (2004). *Amsterdam Short-Term Memory Test*. Amsterdam: Hogrefe Publishing bv.
- Schmand, B., Deelman, B. G., Hooijer, Ch., Jonker, C., & Lindeboom, J. (1996). De itemreeks van de cognitieve screening test vergeleken met die van de mini-mental state examination [The items of the cognitive screening test compared to those of the mini-mental state examination]. *Tijdschrift voor Gerontologie en Geriatrie*, *27*, 29-33.
- Schmidt, M. (1996). *Rey Auditory Verbal Learning Test. A handbook*. Los Angeles: Western Psychological Services.
- Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. *The Clinical Neuropsychologist*, *30*, 515-535.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156-163.
- Slick, J. D., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standard for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545-561.
- Suhr, J. A. & Gunstad, J. (2007). Coaching and malingering: A review. In Larrabee, G. J.

- (Ed.) (2007), *Assessment of malingered neuropsychological deficits* (p.p. 287-311).
New York: Oxford University Press Inc.
- Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, New York: Multi-Health Systems Inc.
- UNESCO (2006). *International Standard Classification of Education: ISCED 1997 (re-edition)*. Montreal: UNESCO-UIS.
- Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zeventenzeventig jaar* [Intelligence and age: Research within Dutch participants aged from 12 to 77 years old]. Dissertation. Assen: Van Gorcum.
- Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. A., Crager, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*, 37-48.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III*. New York: Psychological Corporation.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*, 32-35.
- Young, J. C., Roper, B. L., & Arentsen, T. J. (2016). Validity testing and neuropsychology practice in the VA healthcare system: Results from recent practitioner survey. *The Clinical Neuropsychologist, 30*, 497-514.

7

Chapter 7

GENERAL DISCUSSION



GENERAL DISCUSSION

Research into episodic memory dysfunction in Alzheimer's disease (AD) always had a strong focus on impaired consolidation processes. Nowadays, this research has expanded to how different retrieval support impacts memory and learning in AD. One of the major challenges is to investigate to what extent amnesic mild cognitive impairment (a-MCI) patients and AD patients can remember newly learned information under different types of retrieval support. For this purpose, we developed new applications of the Visual Association Test (VAT; Lindeboom & Schmand, 2003), i.e. parallel versions C and D of the VAT (Lindeboom, Schmand, Meyer, & de Jonghe, 2014), and the measures of memory of the Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017; Meyer & de Jonghe, 2019). In addition, the performance validity measures of the VAT-E allowed us to study and detect low performance validity which may be present in any memory assessment, including studies in mild cognitive impairment and dementia.

The first aim of this thesis was to investigate episodic memory functioning from normal aging to AD dementia. Secondly, knowledge gained about what patients cannot actively remember but can recognize was used to develop memory tests that distinguish between normal effort versus underperformance. The first part examined practical issues in measuring episodic memory in AD dementia. The second part examined the episodic memory functioning across the range normal aging to mild AD dementia. The third part examined the usefulness of the VAT-E in detecting low performance validity in different patient samples, including MCI and dementia. In this final chapter we will discuss findings, methodological considerations, conceptual and theoretical implications, clinical implications, and future research.

Practical issues in measuring episodic memory

Free recall tests are typically used to establish memory deficits in patients, but these tests are found wanting in their ability to describe the remaining memory capacities of patients. Our findings show that, as opposed to word-list free recall tests like the Rey Auditory Verbal Learning Test (RAVLT), AD patients are still able to remember incidentally learned visual associations when they are offered specific memory cues.

We compared score distributions of RAVLT immediate free recall with VAT immediate paired associate recall in a group consisting of healthy elderly controls, a-MCI patients, and mild AD patients by means of hierarchical multiple regression analysis in relation to global cognitive impairment (chapter 2). We found that the RAVLT showed no floor effect in healthy elderly controls, but showed a large floor effect in patient groups, with total scores levelling off for 90% of a-MCI patients and 94% of mild AD patients. By contrast, the VAT showed no floor effects in either patient group, as episodic memory decline was measured linear in healthy elderly controls, a-MCI patients, and mild AD patients across the range of global cognitive impairment.

These findings indicate that free recall measures can be clinically useful in the early stages of Alzheimer's disease, but soon after lose the ability to describe the patient's remaining memory capacities. For this, clinicians need tests that capture different memory processes, such as visual associative learning and retrieval support during testing.

Episodic memory in Alzheimer's disease

Specifically in clinical settings, it is interesting to know to what extent AD patients can remember newly learned information under different types of retrieval support. Our findings show that, compared to free recall, a-MCI patients and mild AD patients remember newly learned information better when retrieval support based on incidental learning of visual

associations is provided by means of cueing and even more so by means of recognition. In addition, intensified test guidance is useful in reducing the number of patients that dropout as well as keeping healthy subjects focused on easy tasks.

Retrieval impairment

We compared VAT-E multiple-choice cued recognition with VAT-E paired associate recall and VAT-E free recall within healthy elderly controls, a-MCI patients, and mild AD patients (chapter 3). We found that patients recognized newly learned material far better than when they had to recall this material. While a-MCI patients or mild AD patients remembered only 12 or 8 percent of the learned material on free recall, these patients were able to remember 77 or 65 percent of this material during the recognition phase of the test. In line with this score profile, also healthy elderly controls improved on recognition compared to recall. In addition, chapter 5 showed a similar score profile for the VAT and the RAVLT. Patients with a-MCI and patients with mild AD improved substantially on VAT delayed multiple-choice cued recognition compared to VAT delayed paired associate recall (chapter 5), and also improved substantially on RAVLT delayed recognition compared to RAVLT immediate free recall (chapter 3) or RAVLT delayed free recall (chapter 3 and 5).

These findings indicate that episodic recognition memory is largely preserved in a-MCI patients and mild AD patients when compared to recall. This suggests that rather than impaired consolidation alone, retrieval processes play a large part in the episodic memory functioning of AD patients.

Cueing and recognition

We examined the extent of the floor effects of VAT-E paired associate recall, VAT-E free recall, and VAT-E multiple-choice cued recognition in a group consisting of healthy elderly

controls, a-MCI patients, and mild AD patients in relation to global cognitive impairment (chapter 4). No floor effects were found for healthy elderly controls on either of these measures. For paired associate recall we found a floor effect in 41% of a-MCI patients and 62% of mild AD patients. For free recall we found a floor effect in 73% of a-MCI patients and 84% of mild AD patients. For multiple-choice cued recognition we did not find a floor effect in a-MCI patients or mild AD patients.

These findings indicate that the varying levels of retrieval support provided by the measure of memory of the VAT-E resulted in varying levels of difficulty with free recall, paired associate recall, and multiple-choice cued recognition being most difficult, intermediate and least difficult respectively. This enables clinicians to more accurately determine the level of retrieval support that may be beneficial for patients with memory difficulties.

Intensified test guidance

We compared the number of missing values between intensified test guidance and routine test administration for the VAT and the RAVLT in healthy elderly controls, a-MCI patients, or mild AD patients (chapter 5). For VAT immediate paired associate recall, we found that, for all groups, the all-cause number of missing values under intensified test guidance was only 2.2%, while under routine test administration this was 13.3%. For RAVLT delayed free recall and RAVLT delayed recognition, we found that, for all groups, the all-cause number of missing values in both conditions was similar, which was caused by the high number of missing values in the AD groups in both conditions. This suggests that, despite intensified test guidance, the delayed free recall trial is still far too difficult for these more severely impaired patients (recognition is only administered when recall is completed). In addition, we compared mean test results between intensified test guidance and routine test administration

for VAT immediate paired associate recall, RAVLT delayed free recall, and RAVLT delayed recognition in healthy elderly controls, a-MCI patients, and mild AD patients (chapter 5). We found that, although memory performances in general did not improve, healthy elderly controls performed better under intensified test guidance on RAVLT delayed recognition.

These findings indicate that intensified test guidance is only effective in motivating AD patients not to give up, i.e. reduce dropout rates, if a test is not too difficult for them. In addition, although intensified test guidance does not improve memory performance, it does prevent unnecessary reduction in memory performance due to loss of focus when tests are very easy for subjects, such as recognition is for healthy elderly controls.

Performance validity of episodic memory

Test result of memory assessments may be distorted by patients who show low performance validity. Our findings show that the performance validity measures of the VAT-E are highly effective in differentiating a-MCI patients and mild AD patients from persons who were instructed to feign memory deficit, and have adequate sensitivity and specificity for Malingering of Neurocognitive Dysfunction in litigating patients.

We compared the total scores of the VAT-E performance validity measures of healthy elderly controls, a-MCI patients, and mild AD patients with persons who were instructed to feign memory deficit by calculating sensitivity and specificity for a range of cut-off scores (chapter 6 study 1). For each patient group, we selected optimal cut-off scores based on the maximum value of Youden's J statistic. We found that the VAT-E differentiated a-MCI patients with a specificity of 93 to 100% and mild AD patients with a specificity of 92 to 100% from persons who were instructed to feign memory deficit with a sensitivity of 86 to 100%. In addition, based on the optimal cut-off scores of the a-MCI group, we found that the VAT-E was in perfect agreement with the Test of Memory Malingering in classifying healthy

elderly controls and persons who were instructed to feign memory deficits. Also, we used the optimal cut-off scores based on the a-MCI group to investigate whether these cut-off scores were able to differentiate litigating patients classified as Malingering of Neurocognitive Dysfunction from litigating patients who were classified as non-Malingering of Neurocognitive Dysfunction (chapter 6 study 2). We found that the VAT-E differentiated Malingering of Neurocognitive Dysfunction with a sensitivity of 54% from non-Malingering of Neurocognitive Dysfunction with a specificity of 97%. In addition, based on the optimal cut-off scores of the a-MCI group, we found that the VAT-E was in moderate agreement with the Word Memory Test in classifying Malingering of Neurocognitive Dysfunction in a group of litigation patients.

These findings indicate that the VAT-E may be a useful performance validity test, as it adequately differentiates patients with genuine memory impairment from litigating patients who were classified as Malingering of Neurocognitive Dysfunction.

Methodological considerations

Strengths

Our research has several strengths. First, the use of multiple measures of memory in a cross-sectional design enabled us to explore episodic memory impairment along the continuum of the early stages of AD. Second, to the best of our knowledge, we used a new statistical application to analyze floor effects (chapter 2 and 4). That is, we used hierarchical multiple regression to determine if the relationship between global cognitive impairment and memory performance could be explained by a linear or quadratic model. If the quadratic trend significantly increased predictive capacity after the linear trend had been taken into account, this was indicative of a floor effect. Third, the use of a simulation design as well as a known-

groups comparison (chapter 6 study 1 and 2 respectively) enabled us to infer a causal conclusion as well as generalize our results respectively.

Limitations

Some limitations of our research must be mentioned as well. First, we used a cross-sectional design and therefore did not measure episodic memory performance related to support during encoding and retrieval longitudinally. Second, some of our a-MCI patients may not convert to AD dementia, and as such not be a part of the continuum of AD. Third, we only investigated a-MCI or mild AD patients, and thus do not know the episodic memory functioning for the whole range of neurodegenerative cognitive decline starting from normal age-related decline to a-MCI to mild AD to moderate or severe AD. Fourth, in chapter 5 we did not know for which cases tests were not administered or stopped during testing, thus we could only report all-cause missing values, and not specify why they were missing. Fifth, the optimal cut-off scores of the VAT-E performance validity measures established in the simulation design may not be generalizable to real-life, as the behavior elicited by the feigning instruction may not have mimicked the behavior in the real world well enough, because the subjects who were instructed to feign memory deficit lacked external incentive, and the € 10,- reward may not have been sufficient motivation. At last, the classification Malingering of Neurocognitive Dysfunction was based on psychometric indicators and therefore may not actually represent persons who show low performance validity. Although the use of psychometric indicators is the accepted approach for identifying low performance validity (Bush et al., 2005; Guilmette, 2013; Heilbronner et al., 2009), the most evident indicator of low performance validity would be a patient admitting he did not put enough effort into his test taking. However, patients showing negative response bias are rarely inclined to do so.

Conceptual and theoretical implications

The general idea that the episodic memory impairment of AD patients is characterized by impaired consolidation alone (Cummings, Dubois, Molinuevo, & Scheltens, 2013; Dubois et al., 2010; Dubois et al., 2007; Dubois et al., 2014; Dubois et al., 2016) is too narrow and may need to be extended by our findings that type of retrieval processes and retrieval support also moderate the level of memory performance in AD patients (chapter 2-4). Although previous studies reported decreased sensitivity to semantic cueing in AD patients compared to healthy elderly controls (Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994; Tounsi et al., 1999) or non-AD disorders such as progressive supranuclear palsy (Pillon et al., 1994), the within group results of these studies also suggest that, within group, AD patients improve on semantic cueing. In line with our findings, a meta-analysis from Bäckman et al. (2005) showed that recognition memory is more intact than free recall in a-MCI patients and mild AD patients, and mouse models of early Alzheimer showed preserved recognition memory (Roy et al., 2016).

Findings in this thesis may have theoretical implications regarding the role of the medial temporal lobe in memory. The prevailing view being that the structures of this lobe, i.e. hippocampal region, entorhinal cortex, perirhinal cortex, and parahippocampal cortex (Squire, Stark, & Clark, 2004), contribute to episodic memory in a similar way (Squire & Bayley, 2007). However, although our a-MCI patients and mild AD patients all have medial temporal lobe atrophy supportive for probable AD (Scheltens et al., 1992), episodic memory performance improves on recognition compared to recall in these patients (chapter 3). Since medial temporal lobe atrophy is defined as volume loss in the hippocampal region and the entorhinal cortex (Dubois et al., 2007), this suggests that the other medial temporal lobe structures have a different function. In line with our findings, Montaldi and Mayes (2010) postulate that the perirhinal cortex and the parahippocampal cortex are essential for visual

episodic recognition memory, whereas the hippocampal region is essential for recall, and that the perirhinal cortex is the mediating structure to subsequently increase activity in associated cortices such as the frontal cortex, the cingulate cortex, the parietal cortex, and the thalamus (Montaldi et al., 2006).

Visual associative learning integrates existing ways of thinking about encoding into a new encoding paradigm, which enhances the engagement with the test material and thus establishes more elaborate, longer lasting, and stronger memory traces (Craik & Lockhart, 1972). First, as stated by dual coding theory (Paivio, 1991) pictures are remembered better because they are encoded visually as well as verbally. Second, the visual associations depict an interaction, which establishes a strong link between the cue and the target (Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002). Third, because these interactions are unexpected they are elaborated on more than expected interactions (Birngruber, Schröter, & Ulrich, 2014; Ulrich, Nitschke, & Rammsayer, 2006).

Clinical implications

Our findings have relevance for diagnostic assessment, as reduced benefit from recognition testing after recall may not be indicative of AD, but rather of the severity of the episodic memory impairment. Current core diagnostic criteria for probable AD state that the episodic memory impairment consists of memory performance that does not improve with recognition testing after a patient showed a deficit on recall (Cummings et al., 2013; Dubois et al., 2010; Dubois et al., 2007; Dubois et al., 2014; Dubois et al., 2016). However, our findings show that even preserved episodic recognition memory may be compatible with AD being the correct diagnosis (chapter 3).

The determination of the level of difficulty of the measures of memory of the VAT-E (chapter 4) enables clinicians not only to establish an episodic memory impairment, but also

provides him with helpful information regarding how to support memory functioning in memory disturbed patients. First, clinicians can establish an impairment based on the norms for free recall (the most difficult measure) (Meyer & de Jonghe, 2019). Next, by inspecting the test scores of paired associate recall (the intermediate difficult measure) and multiple-choice cued recognition (the least difficult measure) clinicians can determine the remaining memory capacities. If a patient improves on paired associate recall and to a large extent on multiple-choice cued recognition, this suggests that the patient is able to benefit more from retrieval support than when a patient does not improve on paired associate recall and only to a small extent on multiple-choice cued recognition. In this way, clinicians can better align health care with patients' memory needs. If the remaining memory capacities are known, clinicians can provide health care professionals with tools to improve the memory performance of patients in their daily life. For example, a patient might learn to associate activities of daily living with icons, which subsequently could be used as a cue in his daily surroundings. Also, clinicians can provide health care professionals with tools to better communicate with patients, such as aligning the communication style with the retrieval support a patient needs, i.e. using cues or multiple-choice in conversation.

Our findings have implications for clinical trials and for monitoring memory impaired patients. Sensitive measures of change, i.e. measures with no restriction of range at the lower end of the measurement range, such as the VAT and the VAT-E multiple choice cued recognition (chapter 2 and 4) may be of benefit for clinical trials in which effects of intervention programs on episodic memory performance in AD have to be evaluated, and for assessing whether patients convert from a-MCI to mild AD. Also, when measuring change, intensified test guidance can be useful in reducing dropout rates in AD patients.

Since the VAT-E employs an embedded performance validity measure it is less vulnerable to coaching. Knowledge about performance validity tests is readily accessible on

the Internet (Bauer & McCaffrey, 2006), and many litigation attorneys inform their clients about validity scales and how they work (Suhr & Gunstad, 2007). However, stand-alone performance validity tests are more vulnerable to coaching (Suhr & Gunstad, 2007) than measures that use patterns of performance on standard neuropsychological measures (Suhr & Gunstad, 2007) such as the VAT-E embedded measure free recall compared to multiple-choice cued recognition.

Recommendations for future research

In the past decade, considerable progression has been made in the field of episodic memory functioning in AD. However, more research is needed. First, episodic memory performance related to support during encoding and retrieval should be measured longitudinally and in more advanced stages of AD. This will enable inferences about the extent to which patients that progress in disease severity can still benefit from support during encoding and retrieval, and will enable a more detailed differentiation between disease stages. Preferably, the healthy elderly controls and the patients from our studies should be retested in order to obtain follow-up data which could then be supplemented with results from patients with more advanced disease stages. We would expect a gradual loss of the effect of cueing or recognition as AD stages become more severe. Second, our research in chapter 4 enabled us to establish the level of difficulty of the measures of memory of the VAT-E. Since neuropsychological examinations need to be attuned to the level of cognitive abilities of a patient in order to make informed inferences about the remaining memory abilities, the level of difficulty should be determined for other neuropsychological measures as well. For this purpose, our new statistical application (chapter 2 and 4) could be used to investigate floor effects of such measures in patient groups, with larger floor effects being indicative of more difficult measures. Third, it should be investigated whether, besides AD dementia, other types of

dementia also benefit from the visual associative learning paradigm. Typically, cognitive deficits in AD patients are measured with tests of free recall. Also, impaired free recall is often contrasted with intact recognition in the clinical differential diagnosis of AD and other types of disorders, particularly vascular dementia, frontotemporal dementia, Lewy body dementia, and Parkinson's disease dementia. We would expect a gradual loss of the effect of cueing or recognition in other types of dementia similar to AD patients when patients are equated to each other based on the severity of the episodic memory impairment. Fourth, the effect of support during encoding, i.e. visual associative learning, should be separated from the effect of support during retrieval, i.e. cueing or recognition. Memory performance on free recall, cueing, and recognition based on side-by-side learning of the VAT pictures should be compared with visual associative learning. We would expect that the support the visual associations give during encoding are of benefit for performance. Fifth, the effectiveness of cueing or recognition as a tool to improve memory functioning of patients in daily life should be investigated. We would expect such memory performance to improve, as, like in the VAT, in daily life knowledge is often linked with the environment. At last, to enable repeated measurements, new parallel versions of the VAT, i.e. versions E, F, G and H, and a parallel versions of the VAT-E should be developed and validated. The construction of new items should be meticulously calibrated to the original items in accordance with the research that constructed the VAT parallel versions C and D (Meyer, Spaan, Lindeboom, Schmand, & de Jonghe, 2013).

References

- Bäckman, L., Jones, S., Berger, A., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology, 19*, 520-531.

- Bauer, L. & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the internet: Is test security threatened? *Archives of Clinical Neuropsychology*, *21*, 121-126.
- Birngruber, T., Schröter, H., & Ulrich, R. (2014). Duration perception of visual and auditory oddball stimuli: Does judgement task moderate the temporal oddball effect? *Attention, Perception, & Psychophysics*, *76*, 814-828.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ...Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 419-426.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.
- Cummings, J. L., Dubois, B., Molinueva, J. L., & Scheltens, P. (2013). International work group criteria for the diagnosis of Alzheimer disease. *The Medical clinics of North America*, *97*, 363-368.
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. J., DeKosky, S. T., Barberger-Gateau, P., ...Scheltens, P. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurology*, *9*, 1118-1127.
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ...Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, *6*, 734-746.
- Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinueva, J. L., Blennow, K., ...Scheltens, P. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurology*, *13*, 614-629.
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., ...Jack, C. R. jr. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic

- criteria. *Alzheimer's & Dementia*, 12, 292-323.
- Guilmette, T. J. (2013). The role of clinical judgement in symptom validity assessment. In Carone, D. A. & Busch S. S. (Eds.) (2013), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p.p. 31-43). New York: Springer Publishing Company, LLC.
- Heilbronner, R. L., Sweet, J. L., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093-1129.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test. Manual*. Leiden: PITS bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73, 126-133.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual]. Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2019). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual] (2nd ed.). Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A., Spaan, P. E. J., Lindeboom, J., Schmand, B., & de Jonghe, J. F. M. (2013). *Ontwikkeling en validatie van twee parallelvormen met herkenning en uitgestelde reproductie voor de visuele associatie test (VAT). Bruikbaar bij differentiaal diagnostiek Alzheimer dementie en monitoring van beloop*. [Development and validation of two parallel forms including recognition and delayed recall for the Visual

- Association Test (VAT). Usable in differential diagnosis of Alzheimer's disease and monitoring the course of the disease]. Master's thesis. Amsterdam: Universiteit van Amsterdam.
- Montaldi, D. & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*, *20*, 1291-1314.
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*, 504-520.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255-287.
- Petersen, R. C., Smith, G. E., Ivnik, R. J., Kokmen, E., & Tangalos, E. G. (1994). Memory function in very early Alzheimer's disease. *Neurology*, *44*, 867-872.
- Pillon, B., Deweer, B., Michon, A., Malapani, C., Agid, Y., & Dubois, B. (1994). Are explicit memory disorders of progressive supranuclear palsy related to damage to striatofrontal circuits? Comparison with Alzheimer's, Parkinson's and Huntington's diseases. *Neurology*, *44*, 1264-1270.
- Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature*, *531*, 508-512.
- Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H.C., Vermersch, P., Kuiper, M., ... Kuiper, M. (1992). Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*, 967-972.
- Squire, L. R. & Bayley P. J. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology*, *17*, 185-196.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual*

Review of Neuroscience, 27, 279-306.

Suhr, J. A. & Gunstad, J. (2007). Coaching and malingering: A review. In Larrabee, G. J. (Ed.) (2007), *Assessment of malingered neuropsychological deficits* (p.p. 287-311). New York: Oxford University Press Inc.

Tounsi, H., Deweer, B., Ergis, A., Linden, M. van der, Pillon, B., Michon, A., et al. (1999). Sensitivity to semantic cueing: An index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Disease and Associated Disorders, 13, 38-46.*

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research, 70, 77-87.*

A

ADDENDUM

SUMMARY

SAMENVATTING (DUTCH SUMMARY)

KNOWLEDGE VALORIZATION

AUTHOR AFFILIATIONS

PUBLICATIONS

DANKWOORD (ACKNOWLEDGEMENTS)

CURRICULUM VITAE

SUMMARY

Research into episodic memory dysfunction in Alzheimer's disease (AD) always had a strong focus on impaired consolidation processes. Nowadays, this research has expanded to how different retrieval support impacts memory and learning in AD. One of the major challenges is to investigate to what extent amnesic mild cognitive impairment (a-MCI) patients and AD patients can remember newly learned information under different types of retrieval support. For this purpose, we developed new applications of the Visual Association Test (VAT), i.e. parallel versions C and D of the VAT, the measures of memory of the Visual Association Test-Extended (VAT-E), and the performance validity measures of the VAT-E.

This thesis consists of three parts. The first part (chapter 2) examined practical issues in measuring episodic memory in AD dementia. The second part (chapter 3-5) examined the episodic memory functioning across the range normal aging to mild AD dementia. The third part (study 1 and 2 of chapter 6) examined the usefulness of the VAT-E in detecting low performance validity in different patient samples, including MCI and dementia.

In **chapter 2** we investigated whether the Rey Auditory Verbal Learning Test (RAVLT) or the VAT can be used to monitor the progression of the episodic memory impairment in a- MCI and AD. In a group consisting of healthy elderly controls, a-MCI patients, and AD patients, floor effects of RAVLT immediate free recall and VAT paired associate recall were compared in relation to global cognitive impairment. Our findings show that, as opposed to word-list free recall tests like the RAVLT, AD patients are still able to remember incidentally learned visual associations when they are offered specific memory cues.

In **chapter 3** we investigated whether the episodic memory impairment in AD can be characterized as a consolidation or a retrieval impairment. The VAT-E measures of memory paired associate recall and free recall were compared with multiple-choice cued recognition

within groups of a-MCI patients and mild AD patients. Our findings showed that episodic recognition memory is largely preserved in a-MCI patients and mild AD patients when compared to recall. This suggests that rather than impaired consolidation alone, retrieval processes play a large part in the episodic memory functioning of AD patients.

In **chapter 4** we investigated to what extent patients with moderate to severe memory deficits can still benefit from retrieval support such as cueing or recognition when incidentally learning visual associations. In a group consisting of healthy elderly controls, a-MCI patients, and mild AD patients, score distributions of the measures of memory of the VAT-E, i.e. paired associate recall, free recall and multiple-choice cued recognition, were compared in relation to the global cognitive impairment. Our findings showed that the varying levels of retrieval support provided by the measure of memory of the VAT-E resulted in varying levels of difficulty with free recall, paired associate recall, and multiple-choice cued recognition being most difficult, intermediate and least difficult respectively. This enables clinicians to more accurately determine the level of retrieval support that may be beneficial for patients with memory difficulties.

In **chapter 5** we investigated whether intensified test guidance on the VAT or the RAVLT can be used to reduce the number of missing values, and can be used to improve the memory performance of patients. The number of missing values and mean test results on the VAT and the RAVLT were compared between intensified test guidance and routine test administration in groups of healthy elderly controls, a-MCI patients, and mild AD patients. Our findings showed that intensified test guidance is only effective in motivating AD patients not to give up, i.e. reduce dropout rates, if a test is not too difficult for them. In addition, although intensified test guidance does not improve memory performance, it does prevent unnecessary reduction in memory performance due to loss of focus when tests are very easy for subjects, such as recognition is for healthy elderly controls.

In **chapter 6** two studies were presented. The first study established the cut-off scores for VAT-E performance validity measures. In an experimental design, persons who were instructed to feign memory deficit were compared with healthy controls, a-MCI patients, and mild AD patients. The second study investigated the usefulness of these cut-off scores to differentiate litigating patients who show low performance validity from those who do not. In a known-groups comparison, litigating patients classified as Malingering of Neurocognitive Dysfunction were compared with litigating patients classified as non-Malingering of Neurocognitive Dysfunction. Our findings show that the performance validity measures of the VAT-E are highly effective in differentiating a-MCI patients and mild AD patients from persons who were instructed to feign memory deficit, and have adequate sensitivity and specificity for Malingering of Neurocognitive Dysfunction in litigating patients. These findings indicate that the VAT-E may be a useful performance validity test, as it adequately differentiates patients with genuine memory impairment from litigating patients who were classified as Malingering of Neurocognitive Dysfunction.

In **chapter 7** we discussed findings, methodological considerations, conceptual and theoretical implications, clinical implications, and future research. There are three main implications of our research. First, the general idea that the episodic memory impairment of AD patients is characterized by impaired consolidation alone is too narrow and may need to be extended by our findings that type of retrieval processes and retrieval support also moderate the level of memory performance in AD patients (chapter 2-4). Second, our findings have relevance for diagnostic assessment, as reduced benefit from recognition testing after recall may not be indicative of AD, but rather of the severity of the episodic memory impairment (chapter 3). Third, our findings have implications for clinical trials in which sensitive measures of change such as the VAT and the VAT-E multiple choice cued recognition (chapter 2 and 4) may be of benefit to establish effects of intervention programs

Addendum

on episodic memory performance in AD, and for assessing whether patients convert from a-MCI to mild AD.

SAMENVATTING (DUTCH SUMMARY)

Onderzoek naar het disfunctioneren van het episodisch geheugen in de ziekte van Alzheimer (AD) is altijd sterk gefocust geweest op gestoorde consolidatie processen. Heden ten dage heeft dit onderzoek zich uitgebreid naar hoe verschillende vormen van ondersteuning bij het ophalen van de informatie impact hebben op geheugen en leren in AD. Eén van de grootste uitdagingen is om te onderzoeken in welke mate patiënten met een amnestische milde cognitieve stoornis (a-MCI) en patiënten met AD zich nieuw geleerde informatie kunnen herinneren onder verschillende vormen van ondersteuning bij het ophalen van informatie.

Met dit doel ontwikkelden we nieuwe toepassingen van de Visuele Associatie Test (VAT), i.e. parallel versies C en D van de VAT, de geheugenschalen van de Visuele Associatie Test-Extended (VAT-E) en de symptoomvaliditeitsschalen van de VAT-E.

Deze dissertatie bestaat uit drie delen. Het eerste deel (hoofdstuk 2) onderzoekt praktische aspecten van het meten van het episodisch geheugen in AD dementie. Het tweede deel (hoofdstukken 3-5) onderzoekt het functioneren van het episodisch geheugen langs de range van normale veroudering tot milde AD dementie. Het derde deel (studie 1 en 2 van hoofdstuk 6) onderzoekt de bruikbaarheid van de VAT-E voor het detecteren van onderpresteren in verschillende groepen patiënten, inclusief MCI en dementie.

In **hoofdstuk 2** onderzochten we of de Rey Auditory Verbal Learning Test (RAVLT) of de VAT bruikbaar zijn voor het monitoren van de progressie van de episodische geheugenstoornis in a-MCI en AD. In een groep die bestond uit gezonde oudere controles, a-MCI patiënten en AD patiënten werden bodem effecten voor RAVLT onmiddellijke vrije reproductie en VAT gepaarde associatie vergeleken in relatie tot de globale cognitieve stoornis. Onze bevindingen toonden aan dat, in tegenstelling tot woordlijst vrije reproductie testen zoals de RAVLT, AD patiënten in staat zijn zich incidenteel geleerde visuele associaties te herinneren als hen specifieke geheugen cues worden gegeven.

In **hoofdstuk 3** onderzochten we of de geheugenstoornis in AD wordt gekenmerkt door een inprentings- of een ophaalprobleem. De VAT-E geheugenschalen gepaarde associatie en vrije reproductie werden vergeleken met multiple choice cued herkenning binnen groepen van a-MCI patiënten en milde AD patiënten. Onze bevindingen toonden aan dat het episodisch herkenningsgeheugen grotendeels is behouden in a-MCI patiënten en milde AD patiënten als we het vergeleken met gepaarde associatie en vrije reproductie. Dit suggereert dat niet alleen inprentingsproblemen, maar ook ophaalproblemen een belangrijke rol spelen bij het functioneren van het episodisch geheugen in AD patiënten.

In **hoofdstuk 4** onderzochten we in welke mate patiënten met matige tot ernstige geheugenstoornissen baat hebben bij het krijgen van ondersteuning bij het ophalen van incidenteel geleerde visuele associaties, zoals cues of herkenning. In een groep die bestond uit gezonde controles, a-MCI patiënten en milde AD patiënten werden score distributies van de geheugenschalen van de VAT-E, i.e. gepaarde associatie, vrije reproductie, en multiple choice cued herkenning, vergeleken in relatie tot de globale cognitieve stoornis. Onze bevindingen toonden aan dat de verschillende mate van ondersteuning die gegeven wordt door de geheugenschalen van de VAT-E resulteert in verschillende mate van moeilijkheid met vrije reproductie, gepaarde associatie, en multiple choice cued herkenning zijnde respectievelijk het meest moeilijk, gemiddeld moeilijk, en het minst moeilijk. Hierdoor kunnen clinici meer accuraat bepalen wat de mate van ondersteuning is waar patiënten met geheugenproblemen baat bij hebben.

In **hoofdstuk 5** onderzochten we of geïntensiverde test begeleiding voor de VAT of de RAVLT bruikbaar is om het aantal missing values terug te brengen en of dit bruikbaar is om de geheugenprestatie van patiënten te verbeteren. In groepen gezonde oudere controles, a-MCI patiënten, en milde AD patiënten werden tussen twee condities, i.e. geïntensiverde test begeleiding en standaard test afname, het aantal missing values en de gemiddelde test

resultaten van de VAT en de RAVLT vergeleken. Onze bevindingen toonden aan dat geïntensiverde test begeleiding alleen effectief is in het motiveren van AD patiënten om niet op te geven, i.e. het aantal missing values terug te brengen, als een test niet te moeilijk is voor hen. Daarbij, alhoewel geïntensiverde test begeleiding niet de geheugenprestatie verbetert, voorkomt deze begeleiding wel onnodige vermindering van de geheugenprestatie als gevolg van concentratieverlies als tests heel erg gemakkelijk zijn voor personen, zoals herkenning dat is voor gezonde oudere controles.

In **hoofdstuk 6** worden twee studies besproken. De eerste studie stelt vast wat de afkapwaarden van de VAT-E symptoomvaliditeitsschalen zijn. In een experimentele opzet vergeleken we personen die waren geïnstrueerd om geheugenproblemen te veinzen met gezonde controles, a-MCI patiënten en milde AD patiënten. De tweede studie onderzocht de bruikbaarheid van deze afkapwaarden om onderscheid te maken tussen patiënten die onderpresteren en patiënten die niet onderpresteren. Patiënten die verwickeld waren in een rechtszaak en geclassificeerd waren als Malingering of Neurocognitive Dysfunction werden vergeleken met patiënten die verwickeld waren in een rechtszaak en geclassificeerd waren als non-Malingering of Neurocognitive Dysfunction. Onze bevindingen toonden aan dat de symptoomvaliditeitsschalen van de VAT-E in hoge mate effectief waren in het onderscheiden van a-MCI patiënten en milde AD patiënten van personen die waren geïnstrueerd om geheugenproblemen te veinzen en adequaat waren in de mate van sensitiviteit en specificiteit voor Malingering of Neurocognitive Dysfunction in patiënten die verwickeld waren in een rechtszaak. Deze bevindingen suggereren dat de VAT-E een bruikbare symptoomvaliditeitstest kan zijn, want is adequaat in het onderscheiden van patiënten met een geheugenstoornis van patiënten die verwickeld waren in een rechtszaak en geclassificeerd waren als Malingering of Neurocognitive Dysfunction.

In **hoofdstuk 7** bespreken we onze bevindingen, methodologische overwegingen, conceptuele en theoretische implicaties, klinische implicaties, en toekomstig onderzoek. Er zijn drie belangrijke implicaties die voortkomen uit ons onderzoek. Ten eerste, het algemene idee dat de episodische geheugenstoornis van AD patiënten wordt gekenmerkt door alleen maar gebrekkige consolidatie, is te beperkt en zou uitgebreid moeten worden met onze bevindingen dat het type ophaalproces en type ondersteuning bij het ophalen van geleerde informatie ook de mate van de geheugenstoornis van AD patiënten modereren. Ten tweede, onze bevindingen zijn relevant voor de diagnostiek, want als patiënten verminderd baat hebben bij testen met herkenning dan zou dit wel eens niet indicatief kunnen zijn voor AD, maar veeleer voor de mate van de geheugenstoornis. Ten derde, onze bevindingen brengen implicaties met zich mee voor klinisch onderzoek waar gevoelige meetinstrumenten, zoals de VAT en de VAT-E multiple choice cued herkenning, nuttig kunnen zijn om effecten van interventie programma's op de geheugenprestatie in AD vast te stellen, en voor het meten of patiënten overgaan van a-MCI naar milde AD.

KNOWLEDGE VALORIZATION

The overall aim of our research was to aid clinicians to better align health care with the memory needs of patients and improve the diagnostic process. For this, clinicians need to be able to assess the extent to which patients with amnesic mild cognitive impairment (a-MCI) and patients with Alzheimer's disease (AD) can still benefit from retrieval support across the continuum of episodic memory impairment. To enable this, we developed new applications of the Visual Association Test (VAT; Lindeboom & Schmand, 2003). First, to monitor the progression of the episodic memory impairment, we developed and validated two extra parallel versions C and D of the VAT (Lindeboom, Schmand, Meyer, & de Jonghe, 2014; Meyer, Spaan, Lindeboom, Schmand, & de Jonghe, 2013). Second, to cover a broad range of episodic memory impairment, we developed and validated the measures of memory of the Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017; Meyer & de Jonghe, 2019). Third, to prevent distortion of test results, we developed and validated the performance validity measures of the VAT-E (Meyer & de Jonghe, 2017). In this chapter we will discuss societal relevance, target groups, products, innovation, and planning and realisation.

1. Societal relevance

The VAT and the measures of memory of the VAT-E enable accurate assessment of memory capacities. This is more important than ever as prevalence of dementia currently is around 50 million cases worldwide and estimated to double every 20 years (Prince et al., 2013; World Health Organization, 2017). Of these cases 60 – 70% is effected by AD (World Health Organization, 2017). In addition, a large portion of demented patients may be non-native speakers, which would struggle with word-list learning tests that are based on the native language and may benefit from a memory test that is based on pictures. In the Netherlands it is estimated that more than 2.5 million persons, i.e. approximately 14.5% of the total

population, grew up speaking a second language besides Dutch (CBS, 2019a; KNAW, 2018). Also, in Europe it is estimated that between 3.7 and 17.8% of adults are non-native speakers, and in the United States of America this is 14.4% (OECD, 2018).

The performance validity measures of the VAT-E enable measurement of memory capacities without distortion of test results by patients who show low performance validity. Prevalence of such low performance validity is estimated to be 6.5% for patients referred to memory clinics (Rienstra et al., 2013), and between 22 and 40% for litigating patients (Larrabee, 2007; Mittenberg, Patton, Canyock, & Condit, 2002). In addition, when subjects are low literate this could confound test results on validity scales such as the Word Memory Test (Green, 2003) where word pairs have to be learned by reading. In such cases, validity scales based on pictures, such as the VAT-E, are more suitable for use. In the Netherlands it is estimated that 1.9 million of adult persons are low literate, i.e. approximately 13.9% of the total population (Algemene Rekenkamer, 2016; CBS, 2019b), and, it is estimated that 15.5% of adults are low literate in Europe, the United States of America, Canada, Australia, South-Korea, Japan, and Russia (Grotlüschen, Mallows, Reder, and Sabatini, 2016).

2. Target audience

Elderly patients may benefit from being administered the VAT or the VAT-E as these tests are more suitable for more vulnerable patients than word list learning tests. Submitting elderly patients to demanding memory tests such as word list learning free recall tests may discourage them, leading to a distorted observation of participation, which is important for clinical assessment of behaviour. Also, extensive assessment of elderly patients with too demanding tests may lead to a distorted representation of the actual memory capacities. By contrast, encoding by means the visual associative learning is less burdensome and taxing for elderly patients, since it facilitates learning more than other memory tests, as it is based on the classic

peg word mnemonic, and pictures of unexpected interactions (Birngruber, Schröter, & Ulrich, 2014; Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002; Paivio, 1991; Ulrich, Nitschke, & Rammsayer, 2006). Also, the VAT and the VAT-E are easy to administer and promote incidental learning, and the VAT is short.

Patients with dyslexia and patients that are non-native speakers may benefit from being administered the VAT and the VAT-E, since the visual associative learning paradigm is based on pictures. First, the first three performance validity measures of the VAT-E, i.e. immediate recognition, delayed recognition, and the consistency between immediate and delayed recognition, are language-independent during encoding as well as retrieval, although in case of non-native speakers instructions may need to be given by a translator. Second, the VAT and the VAT-E paired associate recall, free recall, and multiple-choice cued recognition are language-independent during encoding, although in case of non-native speakers retrieval responses as well as instructions may need to be given by a translator in case of VAT-E paired associate recall and free recall.

Clinicians may benefit from the varying levels of retrieval support the VAT-E provides. In general, during neuropsychological assessment, memory deficits in patients are measured by means of free recall, which often is contrasted with recognition. Since the measures of memory of the VAT-E include trials of free recall, cued recall, and recognition, the clinician can see at a glance what the memory score profile is of a patient. In addition, clinicians need to measure episodic memory decline in AD to accurately monitor disease progression. The parallel versions C and D in conjunction with the previous versions A and B enable this. Furthermore, clinicians need new performance validity measures, because when the general public becomes knowledgeable about the existing measures and know how they work they become ineffective. The newly developed VAT-E performance validity measures answer this need.

In research, development of medication to cure AD is a spearhead. Longitudinally designed studies in which clinical trials investigate outcome measure results of AD patients during intervention programs may benefit from the repeated measurements that are enabled by the four parallel versions of the VAT.

Health care professionals and informal caregivers may benefit from VAT and VAT-E test results as they could employ memory tools based on the retrieval support a patient can still benefit from. In this way, daily memory functioning of patients can be improved. For example, if visual associative learning is still adequate in patients, health care professionals could learn patients to associate daily living activities with icons, which subsequently could be used by informal caregivers to plan a patient's day in a day planner or on a planning board. Also, if retrieval support based on recognition is still adequate, health care professionals can train informal caregivers to use multiple-choice in conversation.

3. Products

In 2013, we created 12 new VAT items, i.e. parallel versions of the original paired associate recall VAT versions A and B (Meyer et al., 2013). We carefully calibrated the new items to the original based on experimental psychological research (De Groot & Keijzer, 2000; Snodgrass & Vanderwart, 1980; Van Schagen, Tamsma, Bruggermann, Jackson, & Michon, 1983), i.e. we selected concrete objects of high name agreement, high image agreement, high familiarity, and without an American cultural bias. In this way, we selected 24 black and white line drawings, which were tied together by an interaction in 12 new drawings. In addition, we developed a multiple-choice trial for each item (Meyer et al., 2013), which the original VAT does not have. In this multiple-choice trial the target and three distractors are shown. We selected distractors based on experimental psychological research (De Deyne & Storms, 2018; De Groot, 1980; Van Loon-Vervoorn & van Bakkum, 1991), i.e. high

frequency of word association with the target or the cue. A professional artist drew all pictures in accordance with a priori stipulated guidelines to safeguard similarity with the original (Meyer et al., 2013).

The parallel versions C and D of the VAT were published in 2014 by Hogrefe Publishing B.V. Amsterdam (Lindeboom et al., 2014), which is part of the international Hogrefe Publishing Group. In order to make the VAT available for an international audience the manual was translated and published in three languages, i.e. Dutch, English, and German. Currently, almost 400 VAT versions C and D have been sold.

We developed the Visual Association Test-Extended (VAT-E; Meyer & de Jonghe, 2017; Meyer & de Jonghe, 2019) by joining the items of all four VAT versions together, i.e. versions A, B, C and D, and adding trials of immediate recognition, delayed recognition, consistency between immediate and delayed recognition, and free recall. Also, only the parallel versions C and D have a recognition trial, which we used as multiple-choice cued recognition trial for the VAT-E.

The VAT-E was published in 2017 by Hogrefe Publishing B.V. Amsterdam (Meyer & de Jonghe, 2017). The second edition of the VAT-E was published in 2019 (Meyer & de Jonghe, 2019). In the second edition, we included larger norm groups, and a recent study that used the VAT-E performance validity measures in a group of patients with Korsakoff Amnesia (Oudman et al., 2019). Currently, almost 200 VAT-E tests have been sold.

4. Innovation

The visual associative learning paradigm facilitates encoding as it is based on the classic peg word mnemonic, and pictures of unexpected interactions (Birngruber, Schröter, & Ulrich, 2014; Lindeboom et al., 2002; Paivio, 1991; Ulrich, Nitschke, & Rammsayer, 2006). As far as

we know, currently no such neuropsychological memory test exists in the toolbox of the neuropsychologist.

The VAT enables repeated measurements. Many memory tests are available for clinical neuropsychological assessment. Only few of those tests possess parallel versions, which are needed for repeated measurements. The parallel versions C and D of the VAT answer to this need. Assessing patients more than ones is of importance for diagnosis, monitoring intervention outcomes, and assessing conversion from a-MCI to mild AD to more advanced disease stages.

The performance validity measures of the VAT-E are an addition to the toolbox of the neuropsychologist. As opposed to the frequently used Word Memory Test (Green, 2003), it is based on pictures instead of words. As opposed to the side-by-side presentation of stimuli during testing of other visual-based performance validity tests such as the Test of Memory Malingering (Tombaugh, 1996) and the Non-Verbal Medical Symptom Validity Test (Green, 2008), the VAT-E presents visual associations. Also, the Test of Memory Malingering does not have measures of memory, and the Non-Verbal Medical Symptom Validity Test does not have a multiple-choice measure of memory. In addition, neither test employs the embedded performance validity measure free recall compared to multiple-choice cued recognition.

The measures of memory of the VAT-E are an addition to the toolbox of the neuropsychologist. As opposed to the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987) the VAT-E is based on pictures instead of words. As opposed to the Free and Cued Selective Reminder Test (Buschke, 1984; Grober, Merling, Heimlich, & Lipton, 1997) the VAT-E is based on incidental learning of visual associations instead of explicit learning of side-by-side presented learning material and has an additional recognition trial. Also, learning material of visual-based memory tests often consists of abstract designs, e.g. Wechsler Memory Scale-IV subtest Visual Reproduction (PsychCorp, 2009; Wechsler,

1945), Rey's Complex Figure Test (Osterreith, 1944; Rey, 1941), and Benton Visual Retention Test (Sivan, 1992), which are less applicable to daily life than the pictures the VAT-E uses, and because these abstract designs have to be recalled by drawing, this may confound test results due to visuoconstructive disabilities.

5. Planning and realisation

The VAT versions C and D and the VAT-E first and second edition have already been published (Lindeboom et al., 2014; Meyer & de Jonghe, 2017; Meyer & de Jonghe, 2019). To further validate and extend norm groups for these tests, data collection is still ongoing at the department of geriatric medicine of the Northwest Medical Center Alkmaar the Netherlands, and the neuropsychological practise of clinical neuropsychologist Dr. J.F.M. de Jonghe. There are several opportunities for further development of the VAT and VAT-E. First, development of a parallel version of the VAT-E and four extra parallel versions of the VAT to enable repeated measurement. Second, digitalization of the VAT and the VAT-E to enable ease of administration and reporting score profiles. Third, translation of the manual of the VAT-E in English and German to promote international availability.

References

Algemene Rekenkamer (2016). *Aanpak van laaggeletterdheid*. [Approach to low literacy].

Den Haag: Algemene Rekenkamer.

Birngruber, T., Schröter, H., & Ulrich, R. (2014). Duration perception of visual and auditory oddball stimuli: Does judgement task moderate the temporal oddball effect? *Attention, Perception, & Psychophysics*, *76*, 814-828.

Buschke, H. (1984). Cued recall in amnesia. *Journal of Clinical Neuropsychology*, *6*, 433-440.

- CBS (2019a). StatLine – Bevolking; kerncijfers [StatLine – Population; key figures]. Retrieved October 16, 2019, from <http://opendata.cbs.nl/statline/>.
- CBS (2019b). StatLine – Bevolking; geslacht, leeftijd en burgerlijke [StatLine – Population; gender, age, and marital status]. Retrieved October 11, 2019, from <http://opendata.cbs.nl/statline/>.
- De Deyne, S. & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*, 198-205.
- De Groot, A.M.B. (1980). *Mondelinge woordassociatienormen: 100 woordassociaties op 460 Nederlandse zelfstandig naamwoorden*. [Verbal wordassociation norms: 100 wordassociations for 460 Dutch nouns]. Lisse: Swets & Zeitlinger bv.
- De Groot, A.M.B. & Keijzer, R. (2000). What is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning, 50*, 1-56.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test (CVLT). Adult Version. Manual*. San Antonio, TX: Psychological Corporation.
- Green, P. (2003). *Green's Word Memory Test for Microsoft Windows: User's manual*. Edmonton, Canada: Green's Publishing Inc.
- Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test: User's manual for Microsoft Windows*. Edmonton, Canada: Green's Publishing Inc.
- Grober, E., Merling, A., Heimlich, T., & Lipton, R. B. (1997). Free and cued selective reminding and selective reminding in the elderly. *Journal of Clinical and Experimental Neuropsychology, 19*, 643-654.
- Grotlüschen, A., Mallows, D., Reder, S., & Sabatini, J. (2016). Adults with low proficiency in literacy or numeracy. *OECD Education Working Papers*, No. 131, OECD Publishing Paris.

- KNAW (2018). *Talen voor Nederland*. [Languages for the Netherlands]. Amsterdam, KNAW.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). In Larrabee, G. J. (Ed.) (2007). *Assessment of malingered neuropsychological deficits* (p.p. 34-37). New York: Oxford University Press Inc.
- Lindeboom, J., & Schmand, B. (2003). *Visual Association Test. Manual*. Leiden: PITS bv.
- Lindeboom, J., Schmand, B., Meyer, S. R. A., & de Jonghe, J. F. M. (2014). *Visual Association Test. Manual*. Amsterdam: Hogrefe Publishing bv.
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73, 126-133.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual]. Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A. & de Jonghe, J. F. M. (2019). *Visuele associatietest – Extended. Handleiding*. [Visual Association Test-Extended. Manual] (2nd ed.). Amsterdam: Hogrefe Publishing bv.
- Meyer, S. R. A., Spaan, P. E. J., Lindeboom, J., Schmand, B., & de Jonghe, J. F. M. (2013). *Ontwikkeling en validatie van twee parallelvormen met herkenning en uitgestelde reproductie voor de visuele associatie test (VAT). Bruikbaar bij differentiaal diagnostiek Alzheimer dementie en monitoring van beloop*. [Development and validation of two parallel forms including recognition and delayed recall for the Visual Association Test (VAT). Usable in differential diagnosis of Alzheimer's disease and monitoring the course of the disease]. Master's thesis. Amsterdam: Universiteit van Amsterdam.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of

- malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.
- OECD (2018). *Skills on the move: Migrants in the survey of adult skills*, OECD skills studies, OECD Publishing Paris.
- Osterreith, P. A. (1944). *Le test de copie d'une figure complex* [The Complex Figure Copy Test]. *The Clinical Neuropsychologist*, 7, 9-15.
- Oudman, E., Krooshof, E., van Oort, R., Lloyd, B., Wijnia, J. W., & Postma, A. (2019). Effects of Korsakoff Amnesia on performance and symptom validity testing. *Applied Neuropsychology: Adult*. Advanced online publication.
doi.org/10.1080/23279095.2019.1576180
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255-287.
- Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's & Dementia*, 9, 63-75.
- PsychCorp. (2009). *Wechsler Memory Scale-Fourth Edition (WMS-IV) technical and interpretive manual*. San Antonio, TX: Pearson.
- Rey, A. (1941). *L'examen psychologique dans les cas d'encephalopathie traumatique* [Psychological examination of traumatic encephalopathy]. *Archives de Psychologie*, 28, 286-340.
- Rienstra, A., Groot, P. F. C., Spaan, P. E. J., Majoie, C. B. L. M., Nederveen, A. J., Walstra, G. J. M., ...Schmand, B. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, 35, 59-70.
- Sivan, A. B. (1992). *Benton Visual Retention Test* (5th ed.). San Antonio, TX:

PsychoCorp/Pearson.

Snodgrass, J.G. & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology*, 6, 174-215.

Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, New York: Multi-Health Systems Inc.

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research*, 70, 77-87.

Van Loon-Vervoorn, W.A. & van Bekkum, I.J. (1991). *Woordassociatie lexicon: Gebaseerd op 1299 woorden en 100 proefpersonen*. [Wordassociations lexicon: Based on 1299 words and 100 test subjects]. Lisse: Swets en Zeitlinger bv.

Van Schagen, I., Tamsma, N., Bruggemann, F., Jackson, J.L., & Michon, J.A. (1983). Namen en normen voor plaatjes. [Names and norms for pictures]. *Nederlands Tijdschrift voor de Psychologie*, 38, 236-241.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, 19, 87-95.

World Health Organization (2017). Dementia fact sheet updated December 2017.

Retrieved January 11, 2018, from <http://www.who.int/mediacentre/factsheet/fs362/en/>.

AUTHOR AFFILIATIONS

Leo Boelaarts, MD, MSc

Department of Geriatric Medicine, Northwest Medical Center Alkmaar, Alkmaar, the Netherlands.

Jos F.M. de Jonghe, PhD

Department of Geriatric Medicine, Northwest Medical Center Alkmaar, Alkmaar, the Netherlands.

Jaap Lindeboom, PhD

Retired.

Rudolf W.H.M. Ponds, PhD

Maastricht University, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS).

Philip Scheltens, PhD, MD

Alzheimer Center, Amsterdam University Medical Centers, Amsterdam, the Netherlands.

Ben Schmand, PhD

University of Amsterdam, Amsterdam, the Netherlands and Department of Medical Psychology, Academic Medical Center, Amsterdam, the Netherlands.

Addendum

Pauline E.J. Spaan, PhD

University of Amsterdam, Amsterdam, the Netherlands and Department of Psychiatry and Medical Psychology, Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands.

PUBLICATIONS**International peer reviewed journals**

Meyer, S. R. A., Spaan, P. E. J., Boelaarts, L., Ponds, R. W. H. M., Schmand, B., & de Jonghe, J. F. M. (2016). Visual associations cued recall A paradigm for measuring episodic memory decline in Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 23*, 566-577.

Meyer, S. R. A., Boelaarts, L., Lindeboom, J., de Jonghe, J. F. M., & Ponds, R. W. H. M. (2019). Episodic recognition memory based on incidental learning of visual associations is largely preserved compared to recall in amnesic mild cognitive impairment and mild Alzheimer's disease. *Applied Neuropsychology: Adult*. Advanced online publication. doi:10.1080/23279095.2019.1703705.

Meyer, S. R. A., de Jonghe, J. F. M., Schmand, B., & Ponds, R. W. H. M. (2019). Visual associations to retrieve episodic memory across healthy elderly, mild cognitive impairment, and patients with Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 26*, 447-462.

Boelaarts, L., **Meyer, S. R. A.,** Scheltens, P., & de Jonghe, J. F. M. (2019). Testing episodic memory in elderly subjects: Not as simple as it looks. *Dementia and Geriatric Cognitive Disorders, 9*, 207-216.

Meyer, S. R. A., de Jonghe, J. F. M., Schmand, B., & Ponds, R. W. H. M. (2017). The Visual Association Test-Extended: A cross-sectional study of the performance validity measures. *The Clinical Neuropsychologist, 31*, 789-813.

Neuropsychological tests

Lindeboom, J., Schmand, B., **Meyer, S. R. A.**, & de Jonghe, J. F. M. (2014). *Visual*

Association Test. Manual. Amsterdam: Hogrefe Publishing bv.

Meyer, S. R. A. & de Jonghe, J. F. M. (2017). *Visuele associatietest – Extended*.

Handleiding. [*Visual Association Test-Extended. Manual*]. Amsterdam: Hogrefe Publishing bv.

Meyer, S. R. A. & de Jonghe, J. F. M. (2019). *Visuele associatietest – Extended*.

Handleiding. [*Visual Association Test-Extended. Manual*] (2nd ed.). Amsterdam: Hogrefe Publishing bv.

DANKWOORD (ACKNOWLEDGEMENTS)

Allereerst wil ik dr. Jos de Jonghe bedanken. Dankzij jouw visie zijn de VAT-CD en VAT-E tot stand gekomen. Ik werd steeds weer geïnspireerd door jouw vele verfrissende en nieuwe ideeën en invalshoeken, waaruit bijvoorbeeld jouw geesteskind de VAT-E is ontstaan. Ook en vooral wil ik je bedanken voor de kans die je mij hebt gegeven om vorm en inhoud te geven aan mijn wetenschappelijke ambities. Ik kan mij als de dag van gisteren herinneren dat ik een email van je kreeg waarin je sprak over de mogelijkheid van een promotietraject met daarbij de tekst “Misschien moeten we eens praten over je aspiraties op onderzoeksgebied?”. Dat deed mij veel, want raakte in de kern mijn wens om wetenschappelijk betekenisvol te zijn. Bedankt dat je mij in de jaren die daarop volgden steeds de weg hebt gewezen in wetenschap, neuropsychologie en onderzoek. Ik kijk uit naar het vervolg van onze samenwerking.

In het bijzonder wil ik mijn promotor prof. dr. Rudolf Ponds bedanken. Scherp in mijn geheugen gegrift staat de dag dat ik via het geven van een presentatie op het symptoomvaliditeitscongres in Maastricht solliciteerde om bij jou te mogen promoveren en dat wij elkaar later die dag de hand gaven op die samenwerking. Dat was voor mij een onvergetelijk moment. Bedankt voor het vertrouwen dat je mij hebt gegeven. Je hartelijke, benaderbare en laagdrempelige manier van contact heb ik heel erg gewaardeerd. Daardoor had ik altijd het gevoel met elke vraag bij je terecht te kunnen. Jouw expertise en overzicht op ons vakgebied en de wetenschap in het algemeen, hebben mij steeds geholpen om de rode draad van ons onderzoek vast te houden. Hartelijk dank.

Veel dank gaat uit naar prof. dr. Ben Schmand. De uitgebreide toelichting die je gaf over de theoretische achtergrond van de VAT tijdens een NVN congres en jouw enthousiaste reactie toen Jos en ik voorstelden om parallelversies van de VAT te maken heb ik heel erg

Addendum

gewaardeerd. In het vervolgtraject zijn vooral jouw methodologische en statistische adviezen van onschatbare waarde gebleken. Ook hebben je adviezen om beknopt te schrijven mij zeer geholpen. Dank daarvoor.

Vooraf ook gaat mijn dank uit naar dr. Jaap Lindeboom. De VAT is jouw geesteskind en het was voor mij dan ook heel spannend toen Jos en ik bij jouw thuis kwamen om de tekeningen van de parallelversies aan je voor te leggen. Dank je wel dat je ons het vertrouwen hebt gegeven om de parallelversies te ontwikkelen. Ik voel mij vereerd dat wij dit hebben mogen doen en wil mijn hartelijke dank uitspreken voor de wetenschappelijke begeleiding die je mij hebt gegeven in het traject daarna.

Mijn hartelijk dank gaat uit naar Leo Boelaarts. Bedankt voor je altijd motiverende en steunende woorden. Dat deed mij goed op de momenten dat ik weer eens een major revision voor mijn kiezen kreeg. En vooral ook bijzonder bedankt voor je wetenschappelijke input, waarbij vooral het perspectief van de arts van grote waarde is gebleken.

Bedankt ook dr. Pauline Spaan. Jouw methodologische en vakinhoudelijke adviezen hebben mij enorm geholpen.

De afdeling geriatrie van de Noordwest Ziekenhuisgroep Alkmaar, bedankt voor de mogelijkheid tot het doen van onderzoek.

Dank aan de leden van de beoordelingscommissie, prof. dr. Verhey, prof. dr. Jelicic, prof. dr. Kessels en prof. dr. Scherder, voor het lezen en beoordelen van dit proefschrift.

Alle patiënten en mantelzorgers wil ik graag bedanken voor het meedoen aan het onderzoek.

Dr. Lottte Tavecchio bedankt voor de schrijfvaardigheden die je mij hebt geleerd in de cursus Advanced Academic Writing aan de Universiteit van Amsterdam. Door jouw cursus heb ik geleerd om academische teksten te schrijven. Dat is van onschatbare waarde gebleken, want zonder schrijfvaardigheid geen proefschrift.

In dankbaarheid denk ik terug aan de steun die oma Meyer mij altijd heeft gegeven.

Rinske, Rhea en Lammert, bedankt voor alle ondersteuning in het reilen en zeilen van ons gezin. Oma, oma en opa, jullie zijn een onmisbare steunpilaar.

Mike en Ramon Breeuwer, broers, bedankt voor alle inspirerende, filosofische en tegelijkertijd jolige gesprekken. Dat is steeds een enorme steun geweest en relativerende tegelijk. Dank dat jullie naast mij willen staan tijdens de verdediging.

Bedankt alle vrienden voor de goede gesprekken en de gezelligheid. Dat heeft mij steeds weer nieuwe energie gegeven om door te gaan.

Annemieke, mijn liefde, mijn geluk. Dank je wel voor al je steun. Zonder jou was het niet gelukt. Op alle fronten heb je mij geholpen om het voor elkaar te krijgen. Dat is heel bijzonder en daar ben ik je eeuwig dankbaar voor. Ik verheug mij op de rest van ons leven samen!

Sven en Wessel, ik ben trots jullie vader te zijn. Jullie zijn waar het om gaat!

CURRICULUM VITAE

Sascha Meyer was born in Hamburg, Germany on May 21st 1970. His mother is Dutch and his father is German. He emigrated to the Netherlands when he was 7 and subsequently grew up in Zaandam. He now lives in Haarlem. After graduating from high school he obtained his Bachelor of Business Administration and worked for 11 years as a consultant in recruitment. In 2006, he started to study clinical neuropsychology at the University of Amsterdam combined with his work as a sociotherapist in mental health care. He graduated Cum Laude for his Bachelor of Science in 2011 and Cum Laude for his Master of Science in 2013. For his Master thesis he developed and validated the parallel versions C and D of the Visual Association Test, which were published in 2014. From his graduation onward he worked for three years as a researcher neuropsychology at the Northwest Medical Center Alkmaar, partly on the basis of a research grant, which he received to further develop and validate the Visual Association Test. During this same period, he worked as a psychologist in a neuropsychological practice for examination of litigating patients, as a psychologist in a practice for psychotherapy, and as a psychologist in a nursing home. After this period and until the present day he has been working as a psychologist at the nursing home Kennemerhart in Haarlem. Combined with this work he continued his research and developed and validated the Visual Association Test-Extended, which was published in 2017, as well as started his PhD research at the School for Mental Health and Neuroscience of Maastricht University. In January 2020 he started his clinical training as a general health care psychologist. Sascha is married to Annemieke Wierstra-Meyer. They have two sons, Sven (2008) and Wessel (2011).

