UC/FPCE_2015

Universidade de Coimbra
Faculdade de Psicologia e de Ciências da Educação

Working memory assessment in older adults: Validation and norming studies of the Month Ordering Task

Imke Buekenhout (buekenhoutimke@gmail.com)

Dissertação de Mestrado em Psicologia Clínica e da Saúde, subárea de especialização em Psicogerontologia Clínica, sob a orientação do Professor Doutor José Augusto Simões Gonçalves Leitão.

“And once the storm is over, you won’t remember how you made it through, how you managed to survive. You won’t even be sure, whether the storm is really over. But one thing is certain. When you come out of the storm, you won’t be the same person who walked in. That’s what this storm’s all about.”

Haruki Murakami, *Kafka on the Shore.*

Working memory assessment in older adults: Validation and norming studies of the Month Ordering Task

Abstract: The present work aimed to obtain construct-related evidence for the Month Ordering Task and to examine its reliability (test-retest and internal consistency). We equally sought to establish normative data for use with speakers of European Portuguese. More specifically, we intended to: 1) attain convergent and discriminant validity evidence for the Month Ordering Task, studying the correlations between this task and other working memory tasks (Reading Span Task and Digit Span Backward) and between the Month Ordering Task and measures of constructs less related to working memory (Digit Span Forward, and measures of inhibition and processing speed derived from the Stroop Neuropsychological Screening Test); 2) analyze the Month Ordering Task's reliability; test-retest reliability was investigated by assessing 40 participants on two different occasions (with a 12-14 week interval) and internal consistency was checked by estimating Cronbach's alpha; 3) establish regression-based norms for use with speakers of European Portuguese.
The Month Ordering Task revealed to possess sound psychometric properties, strongly correlating with other measures of working memory and presenting non-significant to moderate correlations with measures of less related constructs (except for the Digit Span Forward, with which a strong correlation was unveiled). A high retest coefficient was obtained, suggesting that the Month Ordering Task is a temporally stable instrument. The internal consistency study revealed that the Month Ordering Task is an internally consistent and homogeneous scale. Linear regression analyses showed that Month Ordering Task performance is influenced by age, gender and years of formal education. In order to control for these influences, a regression-based algorithm was obtained, enabling the user to transform Month Ordering Task scores into standardized Z scores.

Key Words: Month Ordering Task, working memory, older adults, convergent validity, discriminant validity, test-retest reliability, internal consistency, regression-based norms.

Avaliação da memória de trabalho em adultos idosos: Estudos de validação e normativos da Tarefa de Ordenação de Meses

Resumo: Este trabalho teve como objetivo a análise das propriedades psicométricas da Tarefa de Ordenação de Meses e o estabelecimento de normas baseadas na regressão para uso com indivíduos cuja língua materna seja o português europeu. Mais especificamente, os objetivos desta tese são os seguintes: 1) análise da validade convergente e discriminante da Tarefa de Ordenação de Meses, através do estudo das correlações com outros instrumentos de avaliação da memória de trabalho (Tarefa de Amplitude de Leitura e Memória de Dígitos em sentido inverso) e com instrumentos que avaliam construtos menos relacionados com a memória de trabalho (Memória de Dígitos em sentido direto, e medidas de inibição e de velocidade de processamento obtidas através do Teste Stroop Neuropsicológico); 2) estudo da fiabilidade da Tarefa de Ordenação de Meses, nomeadamente a fiabilidade teste-reteste e a consistência interna; de forma a examinar a fiabilidade teste-reteste, uma amostra de 40 participantes foi avaliada duas vezes em momentos distintos (separados por um intervalo de 12-14 semanas), e a consistência interna foi analisada recorrendo ao índice alfa de Cronbach; 3) por último, procuramos obter dados normativos para falantes do português europeu.

Os resultados obtidos sugerem que a Tarefa de Ordenação de Meses possui boas propriedades psicométricas, apresentando fortes correlações com outras medidas de memória de trabalho e correlações fracas a moderadas com instrumentos que visam a avaliação de construtos distintos (com exceção da Memória de Dígitos em sentido direto, tendo sido obtido um coeficiente de correlação elevado). A Tarefa de Ordenação de Meses mostra ter uma boa estabilidade temporal, demonstrando também ser uma escala homogénea com boa consistência interna. Análises de regressão lineares revelaram que o desempenho na Tarefa de Ordenação de Meses é influenciado por diversas variáveis demográficas (idade, género, anos de escolaridade). De forma a poder controlar o impacto destas variáveis demográficas na avaliação da memória de trabalho através do desempenho na Tarefa de Ordenação de Meses, foram obtidas normas baseadas na regressão. Esse algoritmo permite a transformação de resultados brutos em resultados standardizados.

Palavras-chave: Tarefa de Ordenação de Meses, memória de trabalho, adultos idosos, validade convergente, validade discriminante, fiabilidade teste-reteste, consistência interna, normas baseadas na regressão.

Agradecimentos

A realização desta tese nunca teria sido possível sem o apoio fundamental de diversas pessoas e instituições. Correndo o risco de me alongar bastante, gostaria de aqui deixar a minha mais sincera gratidão.

Ao professor Doutor José Augusto Simões Gonçalves Leitão, pela sua paciência infinita, disponibilidade para esclarecer todas as minhas dúvidas (que não foram poucas), transmissão de inúmeros conhecimentos e excelente orientação. Obrigada por me ter deixado trabalhar ao meu próprio ritmo sem nunca me pressionar. Tornou-se, sem dúvida, num exemplo de rigor e de profissionalismo que procuro seguir.

Aos meus pais e à minha irmã Ines. O vosso apoio e paciência foram fundamentais na realização desta tarefa aparentemente interminável. Obrigada por acreditarem em mim e por estarem sempre do meu lado. Sem vocês não estaria onde estou hoje, e espero um dia poder retribuir tudo o que fizeram por mim.

À Carla, a irmã que Coimbra me deu, a melhor amiga de sempre e para sempre. Obrigada por tornares a minha vida infinitamente melhor e mais leve. Agradeço-te por tudo, desde as palavras de apoio e de conforto aos momentos de loucura. Apesar de agora seguirmos caminhos diferentes, sabes que fazemos do longe o perto.

Às minhas colegas de Psicogerontologia, por me terem acolhido tão bem. Gostava de dirigir o meu agradecimento sobretudo à Fabi, à Catarina e à Joana. Fabi, para além de teres sido uma fantástica colega de estágio, revelaste ser uma amiga importante. Obrigada pelas palavras de ânimo, partilhas e demonstrações de amizade. Catarina, as tuas aventuras e desventuras colocam um sorriso no meu rosto e enchem o meu coração. Joana, as nossas cartas sem fim, as discussões filosóficas e literárias e conversas sobre tudo e nada são capazes de melhorar drasticamente o meu dia.

À Anaísa, por também teres desempenhado um papel fundamental no meu percurso académico. Por toda a ajuda que ofereceste, pelo companheirismo, pelas estadias na Figueira.

Ao Rena, pois recebeste-me sempre de braços abertos. Por teres também cuidado da minha vida social, mesmo quando eu não o queria. Por todos os passeios e todas as conversas.

À Jessi, pelo contributo fundamental na pesquisa de material bibliográfico relevante. Por teres sido uma das melhores colegas de casa de sempre, pois transformaste uma simples casa de estudantes num lar.

À Maria, por todos os cafézinhos, sábios conselhos e capacidade de me tranquilizar nos momentos de maior desespero.

Várias pessoas foram fulcrais na procura de participantes. Assim sendo, devo um sincero agradecimento às famílias Neves, Mota, Questroy-Oliveira e Ramos. Receberam-me sempre de forma calorosa, sem nunca pedir nada em troca. A Ana Carvalho e a Helena Paiva foram igualmente fundamentais na procura de cobaias.

Quero também deixar um agradecimento ao Agrupamento de Escolas de Carregal do Sal, em especial ao professor Hermínio Cunha, pela sua disponibilidade e por ter posto a escola à minha disposição, ao professor Fernando Milheiro pela sua genuína preocupação e pelos seus esforços no recrutamento de participantes, e à Doutora Alexandra Gaudêncio, pelo seu entusiasmo e por igualmente ter dado um contributo importante para o completamento da minha amostra.

À PSP de Coimbra, à Universidade Sénior de Carregal do Sal, ao Centro Social Professora Elisa Barros Silva, ao Instituto Universitário Justiça e Paz, à Cáritas Paroquial de Oliveira do Conde e à Associação Portuguesa de Paralisia Cerebral de Viseu (pólo de Oliveira do Conde), pelo interesse neste projeto e por terem colaborado comigo.

Por fim, queria agradecer, do fundo do coração, a todas as pessoas que participaram neste estudo. Obrigada pela vossa paciência e pelo tempo do vosso dia que reservaram especialmente para mim.

Table of Contents

List of Tables

Acronyms List

WM	Working Memory
MOT	Month Ordering Task
RST	Reading Span Task
DS	Digit Span
DSB	Digit Span Backward
DSF	Digit Span Forward
SNST	Stroop Neuropsychological Screening Test
RCWpc-RBWpc	Difference in proportion-correct scores for reading color names presented as (incongruent) colored words (RCWpc) vs. black words (RBWpc)
RCWs-RBWs	Difference in speed while reading color names presented as (incongruent) colored words (RCWs) vs.black words (RBWs)
NCWpc-NCSpc	Difference in proportion-correct scores for naming colors presented as colored words (incongruous hue/color names) (NCWpc) vs. colored squares (NCSpc)
NCWs-NCSs	Difference in speed for naming colors presented as colored words (NCWs) representing incongruous color names vs. colored squares
NCSs	Naming speed of colored squares
NCSsr	Syllable rate (syllables per second) while naming colors presented as colored squares

Introduction

The present work aims to obtain construct-related evidence for the Month Ordering Task, as well as to study its reliability. We equally intend to establish normative data for use with speakers of European Portuguese.

With age, cognitive losses and declines in memory are frequent, with the most commonly affected components being episodic and working memory (Potter & Attix, 2006). There are two forms of cognitive change associated with aging: the so-called benign changes (resulting from nonspecific histopathological brain alterations), and the malignant changes. The latter are considered atypical, and are due to the presence of specific brain pathology (Smith & Rush, 2006). According to Potter and Attix (2006), it is important to distinguish benign changes typical of normal aging from more severe problems related with progressive cognitive disorders. Mild changes that are common with age are often similar to the early signs presented by dementing illnesses. For instance, memory impairment is one of the main symptoms in dementia, whilst it also is one of the most affected domains in healthy aging (Spaan, Raaijmakers, & Jonker, 2005).

Working memory (WM) is thought of as a system that is responsible for the temporary storage and manipulation of information, playing an important role in several cognitive abilities (as, for instance, problem solving, language, among others) (Ricker, AuBuchon, & Cowan, 2010; Shah & Miyake, 1996). Importantly, considering the goals of the present work, WM is one of the most altered memory components in the elderly (compromised by both benign and malignant changes).

WM problems have been detected in illnesses such as Alzheimer’s disease (Calderon et al., 2001; Kensinger, Shearer, Locascio, Growdon, & Corkin, 2003; Baddeley, Logie, Bressi, Della Sala, & Spinnler, 1986), Parkinson’s disease (Gabrieli, Singh, Stebbins, & Goetz, 1996; Kensinger et al., 2003), dementia with Lewy bodies (Calderon et al., 2001), in cases of traumatic brain injury (McAllister, Flashman, Sparling, & Saykin, 2004; McAllister, Flashman, McDonald, & Saykin, 2006) and stroke (Kirshner & Mark, 2009). Accordingly, the study of WM should be included in the neuropsychological assessment of the elderly, as it might indicate the presence of dementia related cognitive disorders, or be relevant in the evaluation of lesions resulting from stroke and traumatic brain injuries, for example. Early (and comprehensive) assessments are important to adequately inform clinical decisions, and in order to be able to provide adequate psychosocial care and support (Spaan et al., 2005).

Despite the importance of WM assessment of older patients, it is common to apply WM measures developed with younger/cognitively healthy participants to older/clinical populations. However, it is not possible to assume the reliability and validity of the WM measures when used with older or clinical populations: empirical evaluation of the candidate tasks/measures to this specific use is required (Wright & Fergadiotis, 2012).

In fact, WM is often assessed with tasks that might be too difficult for older and cognitively impaired individuals. Tasks that present high levels of difficulty may lead to floor effects, with participants scoring at the bottom of the measure. Complex tasks might also be demoralizing, contributing to a diminished effort (Wright & Fergadiotis, 2012). Potter and Attix (2006) refer that cognitive function amidst the elderly is quite heterogeneous: administering tests that are too difficult or too easy considering the patient's functional levels can lead to false negative and false positive errors in the process of diagnosing impairment.

Almor and colleagues' (Almor, Kempler, MacDonald, Andersen, & Tyler, 1999; Almor, MacDonald, Kempler, Andersen, & Tyler, 2001) suggest that WM tasks that are generally used with young participants are too difficult for patients with Alzheimer's disease (AD). The authors sought to explore the difficulty in understanding spoken language presented by mild to moderate AD patients, which is possibly caused by a WM deficit. However, due to the complexity of standardly used WM tasks, the patients could not be adequately assessed. Therefore, Almor and his team developed a new WM task, the Month Ordering Task (MOT), its goal being the evaluation of verbal WM. In this task, participants are asked to put into calendar sequence an increasingly long group of months that is presented to them out of calendar order. This task holds some advantages when compared to other WM instruments (which will be detailed in the theoretical framework, section 4.3.), and its use can be extended to other clinical populations/healthy senior and even younger populations, as has been demonstrated in a number of studies (Goral et al, 2011; Kljajevic, Fratini, Etxaniz, Urdaneta, & Yanguas, 2013). However, for a widespread use of the MOT to be possible, studies exploring the reliability and validity of this measure should be carried out. To our knowledge, such studies have not been implemented yet.

Considering the fact that the most commonly used WM measures might not be suitable for clinical/older populations, and pondering the advantages presented by the MOT, the aim of this study is to obtain construct-related evidence of validity, as well as reliability data for this task. If the MOT reveals to be a valid and reliable instrument, normative data for speakers of European Portuguese will equally be presented. Internal consistency will be assessed through Cronbach's alpha and test-retest reliability will be verified by administering the MOT on two separate occasions (with a testing interval of 12-14 weeks). Construct-related evidence of validity will be attained by interpreting correlations with scores from other instruments. Using linear regression analyses, the contribution of age, education and gender on MOT performance will be determined, so as to establish appropriate normative data for European Portuguese speakers.

The present thesis comprises six different sections. So as to foster the reader's understanding of the complexity of WM and its importance in neuropsychological assessment, a theoretical framework will be provided. Accordingly, the first part of this document will include a comprehensive description of WM, as well as a brief presentation of its neuroanatomical substrates. Two important and influential WM models will be discussed: Baddeley's working memory model and Cowan's embedded processes model.

The relationship between WM and aging will equally be discussed in this first section. We will explore some possible explanations for the nature of this relationship, that insidiously builds-up impairment in older adults (as, for instance, the slowing of processing speed and the inefficiency of inhibitory mechanisms). The pathological processes that may foster WM dysfunction (such as Alzheimer's disease, Parkinson's disease, dementia with Lewy bodies, traumatic brain injury, stroke) are portrayed as well, leading to the discussion of the importance of WM assessment in the elderly. We will further discuss the adequacy of some of the tasks currently used to measure WM performance, presenting what we consider to be their most important limits, while simultaneously explicating the Month Ordering Task, its characteristics and its advantages as a possible tool for WM assessment.

The second section describes the aims of the present study. The third section concerns the methodology of the present study, describing the participants, instruments, procedure, and statistical analyses. The fourth part of this thesis consists of a thorough description of the results that we have obtained (regarding the influence of demographic characteristics on Month Ordering Task performance, test-retest reliability, internal consistency and construct validity), whilst in the last parts the overall discussion and conclusions can be encountered.

I – Theoretical Framework

In this first part of the theoretical framework, we detail the complex nature and neuroanatomical correlates of working memory, its functions and contents. Subsequently, two theoretical working memory models will be explored: Baddeley's working memory model and Cowan's embedded-processes model. These two cadres were chosen due to their influential import in cognitive psychology, representing different viewpoints pertaining to the WM construct still under debate. One model assumes a domain-general framework (Cowan's) whilst the other adopts a partitioned one (Baddeley's). Subsequently, the relation between working memory and aging will be addressed, describing normative as well as pathological age-related changes. In the last section, the importance of working memory assessment will be depicted, mentioning the most widely used measures in different settings (complex span tasks, *n*-back tasks, WM tasks from the WAIS-IV and WMS-IV) and some frequently encountered obstacles in working memory assessment in senior populations. Finally, the Month Ordering Task and its advantages will be discussed.

1. Working Memory

1.1. Definition

Numerous authors have shown interest in the concept of working memory (WM). It has been frequently defined as a workspace or a system that is used for the temporary storage and manipulation of information (e.g. Baddeley, 1998, 2003; Goldman-Rakic, 1996). Strauss, Sherman and Spreen (2006) have described WM as a “limited-capacity store for retaining information over the short-term (seconds to 1-2 minutes) and for performing mental operations on the contents of this store. The contents of WM may originate from sensory inputs but they may also be retrieved from long-term memory” (p. 678). In the vast literature concerning WM, a consensus about the importance of this construct can be found. Several everyday activities rely on the maintenance and manipulation of information. WM is considered to be the “workbench of cognition” (Jarrold & Towse, 2006). This particular memory system seems to be related to several complex cognitive behaviours as, for instance, reading comprehension, reasoning and arithmetic, among others (e.g., Ackerman, Beier, & Boyle, 2005; Kyllonen & Christal, 1990).

Throughout the literature, short-term memory (STM) and WM have often been used interchangeably. However, it is possible to draw a distinction between these two constructs. While STM instantiates the ability to store information over a limited period of time, WM also implies manipulation of the information that has been held in the memory system (Jarrold & Towse, 2006; Luo & Craik, 2008). Nevertheless, WM is not completely separate from STM, as they both involve the storage of information for a brief period of time (Cowan, 2008).

As stated in the definition, the amount of information that can be maintained in WM is limited (Cowan, 2008, 2010; Miller, 1956; Ricker et al., 2010). Miller (1956) proposed that the limit capacity is about seven items (plus or minus two), discussing a stratagem that could be used to increase memory span: chunking. Another method referred by Baddeley (1998, 2000, 2003), Cowan (2008) and Ricker and colleagues (2010) is rehearsal. When these techniques are impeded, the number of items that can be recalled is only around three or four units (Cowan, 2008, 2010). It is not yet clear why these capacity limits exist.

1.2. Neuroanatomical Substrates

A vast array of neuroimaging studies argue that WM cannot be confined to a single brain area, but instead arises from interactions occurring within an ample neural network (D’Esposito, 2007; Sander, Lindenberger, & Werkle-Bergner, 2012). It appears that the prefrontal cortex (PFC) is an essential node in the WM network (e.g., D’Esposito, Postle, Jonides, & Smith, 1999; Goldman-Rakic, 1996; Kane & Engle, 2002). Fuster and Alexander (1971) and Goldman-Rakic (1996) have performed experiments in monkeys using delayed-response tasks. Using recordings from single neurons within the PFC, it was possible to detect sustained levels of neuronal firing during the delay period of the task (when the monkeys recalled a stimulus presented at the beginning of a trial). These patterns of activity in the absence of stimuli or response have been observed for as long as 12 to 15 seconds in the prefrontal neurons.

D'Esposito, Postle, Jonides, and Smith (1999) and D'Esposito, Postle, Ballard, and Lease (1999) proposed a subdivision of WM related functions within the PFC. The ventrolateral PFC is possibly the region where information is initially received from posterior association areas, and kept active so as to guide behaviour. The dorsolateral PFC is likely engaged when the information held in WM requires monitoring and manipulation. Another view suggests an alternative division of the PFC, with the dorsolateral portion being responsible for spatial WM, and the ventrolateral area for non-spatial WM. Several studies concluded that spatial tasks tend to recruit more right-hemisphere areas, whereas verbal WM appears to mostly rely on left-hemisphere areas (e.g., Jaiswal, Ray, & Slobounov, 2010; Smith, Jonides, & Koeppe, 1996). However, the debate on the mapping of WM functions onto PFC regions remains unresolved.

Linden (2007) advances that the short-term storage of information may also require the involvement of areas in the parietal lobes. The parietal lobes (particularly the posterior part) also play an important role in WM retrieval.

1.3. Functions and Contents

WM is usually differentiated along two dimensions: functions and contents (Kane et al., 2004; Oberauer, Süss, Schulze, Wilhelm, & Wittmann, 2000). On the functional facet, three categories can be distinguished: simultaneous storage and processing, supervision, and coordination (Oberauer et al., 2000). As already mentioned, WM has the dual function of holding information in an accessible state while simultaneously transforming that material. Supervision concerns the monitoring and control functions carried out by WM, activating relevant representations and inhibiting irrelevant ones. Nonetheless, Oberauer, Süß, Wilhelm, and Wittmann (2003) found that supervision was only frailly related to the other WM functions. Lastly, coordination refers to the organization of information elements into structures, by identifying and representing new relations among those elements.

Regarding the content facet, there is strong disagreement about the unitary versus fractionate nature of the construct (e.g., Conway et al., 2005; Kane et al., 2004; Oberauer et al., 2003; Shah & Miyake, 1996). According to some authors, this particular memory system can be divided into verbal and spatial WM, whereas other investigators assume that WM is domain-general. Both views have been supported by numerous studies. For instance, Smith et al. (1996) conducted a positron emission tomography (PET) study in order to compare the activation patterns for spatial and verbal WM. Results indicated that spatial tasks recruited right-hemisphere regions, whereas verbal tasks relied on areas located in the left-hemisphere. The authors concluded that different neural regions mediate spatial and verbal WM. On the other hand, Kane and colleagues (2004) performed a latent-variable study, obtaining results that support the unitary view of WM. Participants completed verbal and visuospatial span tasks, and the underlying constructs measured by these tasks appeared to share 70% - 80% of their variance. The study revealed that WM span measures mainly reflect a domain-general system.

2. Models of WM

2.1. Baddeley's WM Model

Baddeley and Hitch proposed, in 1974, a tripartite model of WM. A supervising system dubbed "central executive" was the main component, which, in turn, was assisted by two subsidiary storage systems: the phonological loop and the visuospatial sketchpad.

The central executive is probably the least studied component. It is thought of as a limited-capacity attention control system that is crucial for strategy selection and control, and coordination of cognitive processes, required by complex cognitive activities (Baddeley, 2010; Strauss et al., 2006). According to Smith and Kosslyn (2007), the central executive also decides when information is lodged in the storage systems and supplies a mechanism responsible for the transformation and manipulation of the material held in the phonological loop and/or the visuo-spatial sketchpad. In terms of neuroanatomy, the central executive resides in the frontal lobes (Baddeley, 2010). The existence of this component is upheld by studies examining dual-task performance. Participants typically perform two distinct tasks (one verbal and one visuospatial), each task requiring storage of information in WM. It is assumed that managing performance of the two tasks requires some sort of coordination, this being the responsibility of the central executive. Several studies demonstrated that the central executive is impaired in Alzheimer's disease. Baddeley, Baddeley, Bucks, and Wilcock (2001) carried out a study in which a control sample and a patient sample (diagnosed with probable Alzheimer's disease) performed an auditory and a visual task. Firstly, participants carried out each one of the tasks separately. The difficulty of the tasks was adjusted for each participant individually, with both samples performing at the same level. Subsequently, participants performed the two tasks simultaneously. The authors concluded that the control group displayed a normal performance, whereas the patient sample showed a clear impairment in performance. As the tasks' difficulty levels were carefully controlled, and all participants had the same level of single-task accuracy, the outcomes could not be attributed to poor performance on the single-task condition, but rather to an impairment in task coordination.

The phonological loop is arguably the component for which Baddeley's WM model provides the most detailed account. It can be defined as a system responsible for holding speech-based information (Baddeley, 1998, 2000). The phonological loop can be segregated into two subcomponents: a phonological store that holds information in a phonological format for a few seconds before it fades, and an articulatory rehearsal system that prevents the memory traces from decaying, by means of active rehearsal (Baddeley, 1998, 2000, 2003; Smith & Kosslyn, 2007). Paulesu, Frith, and Frackowiak (1993) conducted a PET study in order to explore the neural correlates of the phonological loop. Results suggested that the supramarginal gyrus is crucial to the phonological store, whereas the *pars opercularis* seems to play an important role in the functioning of the articulatory rehearsal system.

The other storage system is termed visuospatial sketchpad, and is thought to be responsible for holding and manipulating visuospatial representations (Baddeley, 1998, 2003). This system is primarily dependent on the right hemisphere. Several researchers fractionate the visuospatial sketchpad into two subsystems: one accountable for maintaining visual representations and the other for spatial ones (Smith & Kosslyn, 2007). Wilson, Óscalaidhe and Goldman-Rakic (1993) observed that, in monkeys, different neurons in the PFC are involved in processing spatial and object visual features. The dorsal region of the PFC seemed to be engaged in spatial processing, while the ventral PFC subserved processing of object features.

A third subsidiary system was added to the model in 2000 by Baddeley in an attempt to account for some of the limitations of the earlier model. This component, the episodic buffer, provides an interface between the storage systems and long-term memory (LTM). It also supplies a temporary store where material can be maintained when the capacity of the other peripheral subsystems is exceeded (Baddeley, 2003; Wright & Fergadiotis, 2012). The episodic buffer can be seen as a limited capacity system that is capable of integrating information from a variety of sources using a multimodal code. It is assumed to be controlled by the central executive, albeit differing from this component, as the episodic buffer is principally concerned with the storage of information. It is equally assumed to play a crucial role in providing and retrieving information to and from LTM (Baddeley, 2000, 2003). Frontal areas are likely to be important for the episodic buffer (Baddeley, 2000). As this fourth component is a somewhat new idea, not many studies have been carried out yet. However, Prabhakaran, Narayanan, Zhao, and Gabrieli (2000) undertook an investigation that yielded results supporting the existence of this new component. Functional imaging was used to identify brain regions that are involved in the sustenance of integrated information of WM. Participants performed spatial and verbal WM tasks, and were asked to maintain both verbal and spatial information in an integrated fashion. This integration relied heavily on the PFC.

2.2. Cowan’s Embedded-Processes Model

Cowan proposes that WM involves information in LTM that is activated above some threshold (Cowan, 1988, 1999; Dehn, 2008). These traces become activated due to external stimuli or spreading activation (Wright & Fergadiotis, 2012). The author suggests that there is a single memory-storage system that subsumes elements at various levels of activation, that storage system being LTM. More specifically, Cowan’s model distinguishes between three pools of information: 1) a pool of inactive LTM structures; 2) the subset of LTM that is currently activated; and 3) a subgroup of activated structures that are the focus of attention (Cowan, 1999; Dehn, 2008). The subset of elements that is in the focus of attention receives the maximum activation and is readily available for processing (Wright & Fergadiotis, 2012). Items in the activated pool quickly move in and out of the focus of attention, this being determined by the demands of the current situation (Dehn, 2008). Activated but unneeded items are inhibited, although they are readily accessible. Therefore, the focus of attention is thought to be responsible for filtering (Cowan, 2005). Maintenance of information can be attained through attentional refreshing. According to Camos, Mora, and Oberauer (2011), “attention serves to maintain memory traces in an active state, because directing attention briefly to a target strengthens the target’s representations in WM” (p. 232).

It is assumed that a limited focus of attention restricts WM retention and processing. The focus of attention is capacity-limited in the sense that it can only handle three to five chunks of information at a time, whilst these limits do not apply to the broader pool of activated LTM (Cowan, 1999; Dehn, 2008). The allocation of attention is controlled by two processes: the attention orienting system and the central executive (Cowan, 1988, 1999). The first mechanism is driven by novel, important, unpredictable or intense enough stimuli. If the stimuli’s features become activated and a critical threshold is reached, attention is recruited and the element becomes fully encoded and realized. An example is the cocktail party effect. Wood and Cowan (1995) presented participants with two auditory stimuli (spoken messages) simultaneously. The subjects were asked to listen to the message spoken into the right ear and to repeat each word preferrably without errors, and to ignore the left-ear sounds. Participants’ names were inserted into the irrelevant channel. Participants who heard their name displayed increased errors and response lags to the words following the name. Conscious identification was accompanied by an observable attention shift. However, habituation of orienting can occur, as stimuli that remain unchanged over a period of time do not elicit awareness. The central executive is described as “(…) the collection of mental processes that can be modified by instructions or incentives” (Cowan, 1999, p. 65). It is a goal-oriented system, requiring the involvement of voluntary and controlled attentional processes. Therefore, the central executive refers to information-processing operations that are under the voluntary control of the subject. An example, given by Cowan (1988), is volition. Volition can be observed through manipulations of task instructions and motivational variables. Morey, Cowan, Morey, and Rouder (2011) demonstrated that, while employing dual-task performance, when the incentive associated with one of the tasks increases, performance on the other task typically decreases. The authors claim that if a common attentional resource is drawn upon for performance in both tasks, then emphasizing one task should result in an improvement of performance of that task whilst a declining accomplishment of the other task might be observed.

In terms of neuroanatomy, activated memory is thought to be mediated by diverse areas of the association cortex. The central executive lies within the frontal lobes, and the parietal lobes are possibly the neural correlates of the focus of attention. Entry of information into the focus of attention occurs via the thalamus. Orienting of attention involves the locus coeruleus, whereas habituation implicates the hippocampus (Cowan, 1999; Ricker et al., 2010).

3. WM and Aging

3.1. WM and Normative Senescent Changes

Cells in all regions of the nervous system are affected by the aging process. Brain cells suffer increased amounts of oxidative stress, altered energy homeostasis, accumulation of damaged proteins and lesions in their nucleic acids (Mattson & Magnus, 2006). Vast neurochemical, neuroanatomical and functional changes occur in the human brain across the lifespan (Sander et al., 2012). These alterations cause several cognitive abilities to decline, such as memory, inhibition, processing speed, executive functions, and so on (Mattay et al., 2006). Nonetheless, not all forms of memory decline in a similar fashion with advancing age. Quite a few studies demonstrate that WM and episodic memory are the most affected systems by senescent processes, whereas semantic memory, STM, procedural memory and priming show little age-related decline (e.g., Hoyer & Verhaeghen, 2006; Wang et al., 2011).

Diverse pathophysiological changes occur in normal aging, influencing various brain structures. Many of the cognitive deficits caused by senescent processes involve PFC dysfunction (Wang et al., 2011). Greenwood (2000) writes that "(…) phylogenetically newer parts of the brain, including prefrontal areas, are particularly subject to insult" (p. 705). It is known that WM is largely mediated by the PFC, and deficits in WM follow PFC dysfunction (Greenwood, 2000). Frontal-striatal systems are principally susceptible to white matter lesions, grey matter losses, and neurotransmitter depletion (Buckner, 2004). Grey matter volume declines are specially pronounced in frontal and hippocampal regions (Schulze et al., 2011). Although grey matter atrophy occurs, imaging studies imply that, with respect to the functions it underpins, grey matter remains viable and quite resistant to the aging process (Moseley, 2002).

It has been observed that frontal white matter is particularly affected by advancing age, undergoing a marked deterioration in frontal regions (Buckner, 2004). Nevertheless, white matter alterations do not occur solely in the PFC. Diffusion tensor imaging studies indicate that, with increasing age, a loss of integrity is observed in white matter fiber tracts throughout the whole brain, albeit more pronouncedly in the PFC (Moseley, 2002; Schulze et al., 2011). Some domains of cognitive functioning appear to be predominantly vulnerable to white matter decay, such as inhibition and processing speed (Borella, Ghisletta, & de Ribaupierre, 2011). Several authors propose that the decline in inhibition and processing speed is essential in explaining age-related differences in older adults in diverse cognitive functions. It is equally thought that the relation between WM and aging is not a direct one, but is rather mediated by inhibition and processing speed (e.g. Borella et al., 2011; Hasher & Zacks, 1988; Hoyer & Verhaeghen, 2006; Luo & Craik, 2008; Salthouse, 1996). These mechanisms will be described in more detail in the following section.

Alterations in dopaminergic neurotransmission are also frequently observed in the aging brain. It is know that dopamine is essential to PFC functions, including WM. Age-related changes in dopamine transmission may contribute to impairment of WM function in old age (Bäckman et al., 2011).

According to Cabeza (2002), the brain responds to age-related changes in anatomy and physiology by reorganizing its functions. Structural and physiological PFC changes may provoke adjustments in information processing networks (Mattay et al., 2006). During cognitive performance, PFC activity tends to adopt a bilateral pattern in older adults, whereas younger adults consistently show lateralized activation (Cabeza, 2002; Reuter-Lorenz, 2002). For instance, Reuter-Lorenz and coworkers (2000) undertook a study in which PET was used to asses verbal and spatial WM in older and younger participants. Results indicated that younger individuals exhibited greater left-hemisphere activation for verbal materials and right-hemisphere activation for spatial information. Older participants recruited left and right frontal regions for both types of task. This bihemispheric activation possibly reflects the existence of compensatory processes, with older adults recruiting additional areas in order to compensate for neural decline (Cabeza, 2002; Mattay et al., 2006). Activation of additional areas might augment task performance, but older individuals experience more difficulty and expend more effort on tasks when compared to younger persons, as appropriate processing units may not be as promptly available (Reuter-Lorenz, 2002; Reuter-Lorenz et al., 2000; Schneider-Garces et al., 2009). This leads to positing a third mechanism that perhaps accounts for the age-related changes in WM: a reduction in processing resources that occurs with advancing age (e.g. Luo & Craik, 2008; Mattay et al., 2006; Reuter-Lorenz, 2002; Schneider-Garces et al., 2009). This mechanism will equally be described more thoroughly in the next section.

3.1.1. WM and Reduced Processing Speed

Aging is accompanied by a slowing of processing speed, which has diverse consequences for higher order cognition, including WM (Hoyer & Verhaeghen, 2006; Luo & Craik, 2008). Salthouse (1996) describes two distinct mechanisms that are responsible for the link between processing speed and cognition: the limited time mechanism and the simultaneity mechanism. The first one implies that relevant cognitive operations are executed too slowly and therefore may not be completed in the available time. The latter one is thought to operate because if processing is too slow not all relevant information will be available when needed. Products of early processing may be lost (or become obsolete) by the time that later processing is completed, leading to impairments of critical operations.

Various authors have repeatedly demonstrated the role of processing speed as a mediator of age differences in WM (and other cognitive functions). Salthouse and Babcock (1991) conducted a study so as to decompose age differences in WM in more fundamental components. The hypothesized components were processing speed, processing efficiency, storage capacity and coordination effectiveness. Significant negative correlations were observed between age and WM measures. The greatest attenuation of age differences in WM measures was found after statistically controlling for measures of processing speed, with age-related variance being reduced to nearly 2%.

3.1.2. WM and Loss of Inhibitory Functions

Hasher and Zacks (1988) claim that older adults reveal WM deficits due to less efficient inhibitory mechanisms. Inhibition serves three primary functions: 1) selecting which activated representations are allowed to enter WM (access function); 2) suppressing representations that are irrelevant (or no longer relevant) to the current goal (delete function); and 3) preventing prevalent but unsuitable responses (restraint function) (Borella, Caretti, & De Beni, 2008; Hasher & Zacks, 1988). Inefficient inhibition results in more irrelevant information in WM, and permits this material to receive richer processing as it will not be as effectively dampened. This results in the prolonged preservation of irrelevant information in WM, disrupting performance (Hasher, Stoltzfus, Zacks, & Rympa, 1991).

The inhibition-deficit hypothesis has been extensively studied with the negative priming paradigm. Participants are presented with a task, and the trial's distractor becomes the target on the subsequential trial. Young adults' reaction time generally increases when the distractor in one trial becomes the target in the next one. A possible explanation for this effect is that the trial's distractor is inhibited, and when this distractor becomes a target participants may be slowed due to residual inhibition. As older subjects present inhibition deficits, a weakening of this effect has been observed (Hasher et al., 1991; Roleau & Belleville, 1996). Hasher and colleagues (1991) have noted this pattern in their study. The authors employed the negative priming paradigm using the Letter Naming Task. As expected, younger participants showed a negative priming effect, being slowed down in the selection of the target due to residual inhibition. Older adults, on the other hand, demonstrated no negative priming, probably because, as they did not efficiently suppress the distractors, no residual inhibition was experienced.

3.1.3. WM and Reduction in Processing Resources

Older and younger adults show different activation patterns, as older individuals show bilateral activation under conditions that boost lateralized activity in younger subjects (Reuter-Lorenz, 2002). Altered activation patterns are consistently observed in the PFC (Cabeza, 2002).

Individuals usually activate more cortical regions as task load increases. Increasing task loads may even lead younger adults to recruit additional areas. Nevertheless, older adults seem to engage these areas at lower load levels than younger subjects, possibly to compensate for cognitive decline (Schneider-Garces et al., 2009). Reuter-Lorenz, Stanczak, and Miller (1999) argue that aging leads to a diminution of information processing resources. These resources can be seen as neural units with limited processing capacity. More complex tasks require more neural units to be recruited in order to meet processing demands. Simpler tasks typically need fewer units (Mattay et al., 2006; Reuter-Lorenz et al., 1999). Several studies suggest that advancing age causes a decline in the capacity of each neural unit. Even at lower task loads, older adults need to mobilize more resources so as to satisfy processing demands (Schneider-Garces et al., 2009).

Mattay and coworkers (2006) carried out a study in which younger and older participants executed a version of the *n*-back task with increasing memory load (1-back, 2-back, 3-back) while blood oxygenation-level dependent fMRI (functional magnetic resonance imaging) was being conducted. The results indicated that the elderly recruited more resource units at lower complexity levels than the young participants. Greater bilateral PFC activation was found in older persons at 1-back. Within their WM capacity (when older perform as well as younger persons), older subjects showed greater PFC activation then younger individuals. However, beyond their capacity (at 2-back and 3-back) elders exhibited less PFC activation than the younger participants. This decline in PFC activity reflected a decrease in WM performance. It seems that older persons attain their capacity limits before young subjects do, as they are unable to keep the PFC effectively activated at higher task loads.

3.2. WM and Pathological Senescent Changes

WM is not only affected by normative senescent changes, but also by pathological processes. Memory impairment might be a consequence of acquired brain injury (caused, for instance, by a stroke or a blow to the head) (Jo et al., 2009; Maruff & Darby, 2006). A stroke is the "sudden appearance of neurological symptoms as a result of the interruption of blood flow" (Kolb & Whishaw, 2009, p. 749), and is often associated with WM impairments (Ballard et al., 2003; Jo et al., 2009). According to Maruff and Darby (2006), the prevalence of cerebrovascular disorders that may lead to the occurrence of a stroke augments with age and is, therefore, highest in persons with ages over 60 years. A traumatic brain injury (TBI) is a lesion to the brain that is a consequence of a blow to the head (Kolb & Whishaw, 2009). One of the main causes of TBI are falls, putting 64+ age groups at increased risk (T. Morris, 2010). The main complaint after TBI is that of WM related problems, as brain regions important to the WM circuitry are vulnerable to injury in TBI. It is well established that the frontal lobes are frequently damaged, notwithstanding the mechanism of injury (McAllister et al., 2004).

WM impairment might also result from neurodegenerative disorders, with dementia being the most common cause in the elderly population. Whatmough (2010) defines dementia as "(…) a condition of persistent decline in multiple mental domains essential to normal daily living" (p. 277). As Alzheimer's disease (AD) is the most common form of dementia, the majority of the existing research has dealt with this specific disorder. Memory impairment is the hallmark of AD. Although LTM problems have been more frequently referred to, several studies have reported that AD can also be responsible for WM deficits (Kensinger et al., 2003; Whatmough, 2010). WM decline in AD has been mainly explored in the light of Baddeley's model, and various researchers consider that the central executive might be impaired (Baddeley et al., 1986; Carlesimo & Oscar-Berman, 1992). AD patients seem to have difficulties in simultaneously performing two concurrent tasks, given that the regulation of different processes is one of the functions of the central executive (Baddeley et al., 1986). The second most prevalent form of dementia is dementia with Lewy bodies (DLB) (Whatmough, 2010). Although studies are still relatively scarce, DLB patients suffer from WM impairments. Given that Lewy bodies pathology affects frontal, temporal, parietal and subcortical structures, it is expected that all WM components are affected (Calderon et al., 2001). Despite the fact that Parkinson's disease (PD) does not always involve dementia, patients develop several cognitive deficits (R. Morris et al., 1988). Diverse memory functions are impaired in PD, particularly WM. It has been observed that PD patients have a dimished WM span (Gabrieli et al., 1996). This deficit might be due to a reduction of dopamine concentration in the PFC (Beato et al., 2008).

4. WM Assessment

WM is thus affected by both normative and pathological aging processes. Such deficits may be a source of distress and disability, and be responsible for a loss of independence and autonomy (Jo et al., 2009). The elderly population is increasing and, consequently, an increment in memory deficits due to normal aging or pathological processes is to be expected (Bastos Leite, Scheltens, & Barkhof, 2004).

Memory problems are generally the early symptom of a neurodegenerative disorder but can also be purely age-associated. The difference is frequently subtle (Potter & Attix, 2006). Cipolotti and Warrington (1995) state that "(…) it is often necessary to differentiate memory failure due to cognitive deterioration rather than benign senescent forgetfulness" (p. 662).

So as to distinguish normal aging from pathological aging and to identify possible neurodegenerative disorders, a thorough neuropsychological evaluation is called for. As WM is frequently impaired, reliable and valid WM measurements should equally be included. A comprehensive neuropsychological assessment is important, as it is valuable in planning the medical and social care of the patient and in monitoring change over time (Cipolotti & Warrington, 1995).

4.1. Frequently Used WM Tasks

4.1.1. Frequently Used WM Tasks for Research Purposes

4.1.1.1. Complex Span Tasks

Complex span tasks are amongst the most frequently used WM measures, and are commonly employed in individual-differences research (Conway et al., 2005; Redick & Lindsey, 2013). These tasks demand information storage and rehearsal, and the simultaneous processing of supplementary information. Participants are asked to remember stimulus sequences while accomplishing a seconday task (Conway et al., 2005; Jarrold & Towse, 2006; Kane, Conway, Miura, & Colflesh, 2007; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009). The three most used measures are the Reading Span Task (RST), Counting Span Task (CST) and Operation Span Task (OST).

Daneman and Carpenter's (1980) RST was the first tool developed in order to tap the storage and processing functions of WM, and became one of the most widely used WM measures. Participants read sentences aloud and are required to remember the sentence-final words. The sentences are presented in sets that increase in size, and word recall occurs at the completion of each set.

Case, Kurland and Goldberg's (1982) CST has also been regularly administered in order to assess WM. This task involves counting, and participants must store the count totals for later recall. In the original task, subjects are presented with a set of cards, one by one. On each card a number of dots can be seen, which the participant is required to count. After the last card has been displayed and removed, the subject is asked to recall the number of dots on each. Set size is progressively augmented.

In Turner and Engle's (1989) OST, participants have to perform simple arithmetic operations (e.g. 4/2 + 1 = 3) which are followed by a to-be-remembered word. Two to five sets are exhibited before recall is prompted.

Despite the fact that complex span measures are the most extensively used assessment tools in cognitive psychology, studies focusing on the psychometric properties of these tasks have yielded mixed results (Beckman, Hollig, & Kuhn, 2007). Nevertheless, it appears that complex span tasks have acceptable convergent validity, since they correlate well with each other and with measures that are thought to require WM (as, for instance, tests of fluid intelligence and reasoning) (Conway et al., 2005). Kane and colleagues (2004) reported correlations between RST, CST and OST that ranged from 0.55 to 0.73. In another study, Kane and coworkers (2007) found that OST and a measure of fluid intelligence (Ravens Advanced Progressive Matrices Test) were significantly correlated (r=0.33). Significant correlations between complex span tasks and measures of higher order cognition were equally obtained by Daneman and Carpenter (1983). The authors found that RST was significantly correlated with performance on probe questions of a reading comprehension task (r=0.60). Participants were divided into three groups (small spans, intermediate spans, large spans), and individuals with small spans answered questions correctly 51% of the time; participants with intermediate spans were correct 61% of the time, whereas large-span readers were right 84% of the time. As such, it seems that WM is related with reading comprehension.

Additionally, complex span tasks present adequate internal consistency values. Kane and colleagues (2004) observed that the alpha coefficients for OST, RST and CST varied between 0.77 and 0.80. Klein and Fiss (1999) reported an alpha coefficient of 0.75 for OST.

Few studies have examined the test-retest reliability of WM measures. For the OST, Klein and Fiss (1999) attained a test-retest correlation of .88 (with a three-month interval). Beckman and colleagues (2007) reported a test-retest correlation of .70 for RST, with a minimum interval of 4.5 months and a maximum interval of 6 months between testing sessions. Kuntsi, Stevenson, Oosterlaan, and Sonuga-Barke (2001) calculated test-retest correlations for the CST in a sample of children with ages between 7.9 and 15.3 years, with a two-week interval. The analysis yielded a 0.67 correlation. Nonetheless, several authors have attained less adequate test-retest reliability values for the RST. For instance, MacDonald, Almor, Henderson, Kempler, and Andersen (2001) reported a test-retest reliability coefficient of 0.54 for RST, with a two-week interval between testings. Waters and Caplan (1996) achieved a test-retest correlation of 0.41 for the same measure (with a three-month interval). The test-retest reliability coefficients for the RST revealed to be less than optimal, hampering the utility of this particular task (Strauss et al., 2006).

Complex span measures seem to have discriminant validity, as "(…) they do not predict performance on tasks that appear to reflect relatively automatic processing" (Conway et al., 2005, p. 778). Kane, Bleckley, Conway, and Engle (2001) demonstrated that high-span and low-span participants performed equally well on the prossacade condition of the antisaccade task.

Lastly, Conway and colleagues (2005) argue that these complex span tasks are no perfect WM measures, as they are not process pure. For example, the RST assesses WM, but probably also verbal skills. OST is also a measure of WM, surely tapping other capacities such as motivation and mathematical abilities.

4.1.1.2. *N*-back Task

Another measure that is assumed to assess WM processes is the *n*-back task, being extensively used in the cognitive neurosciences (Jarrold & Towse, 2006; Schmiedek et al., 2009). This task was first introduced by Kirchner (1958), who described it as a "(…) task involving rapidly, continuously changing information" (p. 352). Participants are presented with a list of stimuli and are required to decide for each stimulus whether it matches the one presented *n* items before. By varying *n*, the memory load can easily be manipulated (Jaeggi, Buschkuehl, Perrig, & Meier 2010; Schmiedek et al., 2009). The task implies the storage of information in correct serial order, and items need to be available in memory for short time intervals. Transformation of the stored information presumably also occurs, as it is not necessary to retain items that appeared *n* items ago. Therefore, the list of relevant information in memory is constantly updated (Jarrold &

Towse, 2006; Redick & Lindsey, 2013). Nevertheless, it appears that the *n*-back task does not depend solely upon WM processes. An array of other abilities is required, such as inhibition, interference resolution, decision making and selection (Jaeggi et al., 2010).

Regardless of the fact that the *n*-back task has become a standard WM tool, its psychometric properties have remained largely unexplored, and studies on this topic have reported somewhat contradictory results. In terms of convergent validity, the *n*-back task seems not to correlate well with complex span tasks. In a meta-analytic study carried out by Redick and Lindsey (2013), results suggested that complex span tasks and the *n*-back task are weakly correlated, with values ranging from -0.07 to 0.41. Similar outcomes were attained by some other studies. To give an example, Roberts and Gibson (2002) found that the correlations between the *n*-back task (a composite *n*-back score was calculated) and complex span tasks were frail. Specifically, the value of the correlation between *n*-back and CST was 0.24, and between *n*-back and RST the value was -0.15. These frail relations between the *n*-back task and complex span tasks might be explained by the fact that memory is presumably not a unitary function. It is possible that different tasks tap separate components of the WM system, resulting in weak correlations between different WM measures. Another difference between the *n*-back task and complex span tasks is whether performance is driven by recall or recognition. Complex span tasks seem to require active recall, as participants must generate the sequence of items presented on a given trial. On the contrary, *n*-back task performance depends upon familiarity- and recognition-based discrimination processes, given that participants are asked to recognize a current item as an item that was presented *n* positions ago. The weak relation between the different task types may be due to the different processes that are required (Jaeggi et al., 2010; Redick & Lindsey, 2013; Roberts & Gibson, 2002; Waters & Caplan, 2003).

On the other hand, Schmiedek and colleagues (2009) found that updating tasks (such as the *n*-back task) assess WM equally well as complex span measures. Using a latent variable analysis, the authors concluded that the complex span factor and an updating factor correlated very highly ($r=0.96$). As is the case with complex span measures, Kane and colleagues (2007) found that *n*-back performance correlates equally well with measures of higher order cognition (as is the case with fluid intelligence). Correlations between a 3-back task and the Ravens Advanced Progressive Matrices Test ranged from 0.18 to 0.48.

Concerning the internal consistency and test-retest reliability of the *n*-back task, studies have once again obtained mixed outcomes. Forns and coworkers (2014) explored the internal consistency of the *n*-back task in a sample of children with ages between 7 and 9 years, yielding an alpha coefficient of 0.80. DeDe, Ricca, Knilans, and Trubl (2014) equally examined the psychometric properties of the *n*-back task (1-back and 2-back). The 1-back condition presented an alpha coefficient of 0.41, whereas the 2-back condition's internal consistency value was 0.61. Hockey and Geffen (2004) investigated the test-retest reliability of the *n*-back task (0-

back to 3-back), with a 1-week interval between test administrations. Moderate test-retest reliability for accuracy was reported across the four levels (ranging from 0.30 to 0.60), whereas reaction times were very reliable (values varying between 0.69 and 0.86). It is possible that the *n*-back task is a more reliable measure of processing speed than of WM.

4.1.2. Frequently Used WM Tasks in Clinical Settings

Clinical psychologists hold the Wechsler Scales in high esteem, and the Wechsler Adult Intelligence Scale's (WAIS) and Wechsler Memory Scale's (WMS) WM indices are the most widely used instruments to assess WM in clinical settings (Hill, 2008). Both instruments were recently revised and are currently in their fourth edition (Wechsler, 2008, 2009). Both scales have excellent psychometric properties, provide comprehensive and recent normative data based upon samples of considerable size, encompassing an ample age range (Groth-Marnat, 2009; Lichtenberger & Kaufman, 2009).

The WAIS-IV only comprises auditory WM subtests, whereas the WMS-IV includes visual WM tasks (Drozdick, Holnack, & Hilsabeck, 2011; Ward, Bergman, & Hebert, 2012). The WAIS-IV WM index involves three subtests: Digit Span, Arithmetic and Letter-Number Sequencing.

The Digit Span (DS) subtest has been extensively used in psychological assessment since the late 1800's, having already been employed by Ebbinghaus and Galton (Hill, 2008). The DS has been included in all the Wechsler Intelligence Scales and has, until its inclusion in the WAIS-IV, remained relatively unchanged. In earlier versions of the WAIS, DS comprised two conditions: Digit Span Forward (DSF) and Digit Span Backward (DSB). In the DSF part of the test, the individual is read a list of numbers (at a rate of one per second), and is required to repeat them in the same order. In the DSB section, the examinee is also presented with a sequence of numbers, and is asked to recall them in reverse order (Hill et al., 2010; Lichtenberger & Kaufman, 2009). Several authors argue that DSF and DSB are quite different tasks (e.g. Hill, 2008; Reynolds, 1997). DSF has been considered a simple span task, as it is thought to simply require storage of information. DSB, on the other hand, is a more demanding task, requiring storage and also manipulation of information (participants have to mentally reorder the string of numbers in order to answer correctly). Hill (2008) states that, for a task that primarily intends to measure WM, DS "seems to have a heavy immediate memory component inherent in its design" (p. 11). Due to its limitations, the WAIS-IV includes a revised version of the DS, placing greater demands on WM. This new version of the DS includes a third task, Digit Span Sequencing (DSS). The examinee is presented with a sequence of numbers and is asked to repeat them in ascending order (Lichtenberger & Kaufman, 2009). DSF measures different aspects of memory than DSB and DSS. DSF is thought to be a good measure of the phonological loop, whilst DSB and DSS require a more extensive involvement of the central executive, as manipulation of information is required (Lichtenberger & Kaufman, 2009).

The Arithmetic (AR) subtest has equally remained a core subtest of all the Wechsler Intelligence Scales (Hill, 2008). In this subtest, the examinee is read arithmetic problems that have to be mentally solved within a certain time limit, and the problems increase in difficulty. The individual is not allowed to use paper and pencil during this subtest, requiring the utilization of WM in order to remember the details of the problem while it is being solved (Hill et al., 2010; Lichtenberger & Kaufman, 2009). Despite the fact that AR is not an assessment of mathematical abilities per se, performance on this subtest might be influenced to some degree by these abilities. Anxiety related to mathematics has been shown to negatively impact the outcomes on AR (Hill, 2008). Performance on AR is also believed to depend upon educational and professional attainment. Due to these drawbacks, AR was equally recently revised so as to become a stronger WM measure. Coalson, Raiford, Saklofske, and Weiss (2010) write that “arithmetic items were revised to decrease demands on verbal comprehension and mathematical knowledge, thus increasing demands on WM” (p. 16).

The Letter-Number Sequencing (LNS) subtest was developed by Gold, Carpenter, Randolph, Goldberg, and Weinberger (1997) in order to evaluate WM in schizophrenic patients. Subjects are read a series of letters and numbers, and are asked to repeat the numbers first in ascending order, and then the letters in alphabetic order (Groth-Marnat, 2009; Strauss et al., 2006). This subtest might be, according to Hill (2008), one of the best WM measures among the Wechsler subtests. Haut, Kuwabara, Leach, and Arias (2000) undertook a PET study to explore LNS performance, and observed that the brain regions activated by this subtest are commonly triggered by other WM measures. As a WM task, LNS presents several advantages, as it does not include a component that merely measures STM (unlike DS), and as no previous knowledge is required (unlike AR). The examinee is solely required to know the alphabet and the numbers 1 through 9 (Hill, 2008).

As the AR and LNS subtests include a processing component by requiring manipulation of information, both tasks place high demands on the central executive.

The WMS-IV includes two new subtests that assess WM: Spatial Addition and Symbol Span. The former WMS-III tasks principally measured visual storage, generally lacking a processing component. As such, new subtests were included in the WMS-IV that tap both storage and processing components of WM (Drozdick et al., 2011).

The Spatial Addition subtest is based on a modified *n*-back paradigm, and the individual is shown a grid (during 5 seconds) with blue and/or red dots on it. The examinee is asked to remember the location of the blue dots and to ignore the red dots. After that, a second grid with blue and/or red dots is shown. The person is then required to add the two visual images together, and, to provide an answer, is given cards with blue, red and white dots on them and also a grid. The examinee ought to place blue dots on the grid on the locations where the blue dots were previously seen on either page, and a white dot must be put on the locations where blue dots appear on both pages (Groth-Marnat, 2009; Drozdick et al., 2011).

In the Symbol Span subtest, novel visual stimuli are employed that are difficult to verbalize and rehearse, keeping the involvement of verbal WM to a minimum. Examinees are presented with a page containing a series of abstract symbols (for 5 seconds), and are then shown a page with correct designs and foils. The subjects must select the correct designs in the correct order (Groth-Marnat, 2009; Drozdick et al., 2011).

Both tasks are thought to assess the visuo-spatial sketchpad (storage component) and also the central executive (manipulation of information) (Spores, 2013).

4.2. WM Assessment and the Elderly

Despite the growth of the elderly population and the concomitant increase in age-related cognitive deficits, challenges in geriatric neuropsychological assessment still exist. As already stated, WM is one of the most affected systems by normal and pathological aging processes. Nonetheless, there seems to be a lack of valid and reliable WM measures for the evaluation of elder patients.

Waters and Caplan (2003) state that the psychometric properties of WM tasks have not been properly explored in the senior portion of the community. Also, very few researchers have studied the correlations between WM tasks in this fraction of the population. Several authors frequently discuss the inexistence of age-appropriate normative data, possibly hampering test interpretation as there is not a good fit between the normative population and the older patients (Busch, Chelune, & Suchy, 2006; Potter & Attix, 2006). Many normative studies are based on rather small samples, with some age groups comprising fewer than 45 participants. The selected individuals might sometimes not fully match the demographic characteristics of the population they are assumed to portray (Potter & Attix, 2006). More representative samples are needed, as variability within test performance tends to increase with age. The shortage of appropriate normative and psychometric data frequently result in inefficient neuropsychological practice, possibly leading to poorly formulated diagnoses and treatment recommendations (Strauss et al., 2006).

Often, researchers employ WM tasks that are adequate for younger adults but may prove to be too difficult for older participants. Many WM measures have been developed using younger samples and might not be adequate for the assessment of older/cognitively impaired adults. Task administration instructions are generally designed for healthy young adults (Carlesimo & Oscar-Berman, 1992; Conway et al., 2005). Complicated WM measures that present complex information can lead to floor effects, perhaps causing an overestimation of pathology in older adults. High levels of difficulty might be frustrating for the patient/participant, surely affecting task performance (Potter & Attix, 2006). Manning and Ducharme (2010) claim that older patients possibly have a limited experience within a test-taking environment and presumably do not fully understand what is asked of them. Additionally, some WM tasks might be considered artificial as they do not use ecologically valid stimuli. These stimuli do not directly relate to the patient’s experience and knowledge, causing a decrease in motivation and in the interest in the task (Harkin, Rutherford, & Kessler, 2011), a problem probably impacting performance more seriously in older persons.

Despite their scarcity, some studies corroborate the inadequacy of some complex WM tasks for the assessment of older persons. The majority of these investigations focus on AD, as it is the most common form of dementia, and employed the RST, as it became the standard method of evaluating verbal WM in experimental studies. MacDonald and colleagues (2001) found that Daneman and Carpenter's RST was not suitable for the assessment of AD patients' WM, having obtained very inconsistent results. According to the authors, this poor performance might arise from the complex task instructions. Participants seem to find it difficult to undertand and to remember to do two things at the same time (reading out loud and remembering sentence-final words). Complicated instructions that require participants to simultaneously perform different actions seem to be an intrinsic part of several WM tasks. In order to efficiently assess WM, the authors felt the need to devise a new WM task, the Digit Ordering Task. Almor and coworkers (1999, 2001) had to tackle a similar issue. They also found the RST too difficult for AD patients, not yielding sufficient variance for meaningful statistical analysis. Almor and his team equally developed a new WM measure, the Month Ordering Task. This task will be depicted in more detail in the ensuing section. Kempler, Almor, Tyler, Andersen, and MacDonald (1998) also employed the Month Ordering and Digit Ordering Tasks in their study, as the RST is not simple enough for demented patients to execute. One might assume that the same problem applies to other complex span tasks, as they share an underlying structure and are administered in a similar manner (Conway et al., 2005).

Lastly, it is known that WM tasks are not process pure, thus evaluating other cognitive capacities (such as inhibition, processing speed, amidst others). These abilities also suffer age-related declines, therefore hindering WM performance. For instance, the *n*-back task requires the participant to inhibit irrelevant information. Inefficient inhibitory mechanisms will impair task performance, as the memory store will be clogged with irrelevant information (Carlesimo & Oscar-Berman, 1992).

4.3. The Month Ordering Task

As stated earlier, the Month Ordering Task (MOT) was created by Almor and colleagues (1999, 2001), in order to obtain a measure that could easily be used with patients with mild to moderate dementia. The MOT requires simultaneous storage and manipulation of verbal information, like other verbal WM measures. The administration procedure was adapted from Daneman and Carpenter's RST, and participants are asked to put into calendar sequence a progressively longer list of months that are presented out of calendar order.

The MOT might prove to be useful in the assessment of older and/or cognitively challenged patients, as it presents an array of advantages when compared to commonly used WM tasks. The MOT is easy and quick to administer and instructions are simple, assuring good compliance. Also, the task is based on an easily understood idea of ordering months, with the stimuli readily relating to the patients' knowledge. The simple instructions and ecologically valid stimuli assure higher levels of motivation and cooperation, fostering the persons' interest in the proposed task. Since the MOT has been successfully employed in some studies involving AD patients, one can presume that the degree of difficulty is adequate for the elderly and demented populations. As such, floor effects are less expected to occur, turning it into a sensitive measure (Almor et al., 1999, 2001; MacDonald et al., 2001). Ceiling effects should also not take place, since the MOT has been successfully applied in studies involving healthy elder and younger participants, yielding meaningful variance also within the latter group. Goral and colleagues (2011) carried out a study that intended to assess the involvement of switching skills and WM capacity in auditory sentence processing in healthy middle-aged and older adults. Three age groups were created: 55-64 years, 65-74 years and 75-88 years. The MOT revealed to be a suitable tool to assess WM across the different age groups. Month ordering span decreased with age, but participants in each age group attained satisfactory results with adequate variability, and no floor or ceiling effects were observed. The authors also obtained outcomes that supported the predictive power of the MOT. WM span (as measured by the MOT) predicted performance on the sentence tasks, and individuals with higher WM spans performed at higher accuracy levels than those with lower WM spans. Kljajevic and coworkers (2013) also undertook a study in which the MOT was employed as a measure of verbal WM. The sample consisted of 39 healthy participants with different ages: 18 were younger persons and 21 were older. The group of older participants was divided into two subgroups, according to their MoCA scores: the normal MoCA (NM) scores group (≥26) and the low MoCA (LM) scores group (<26). Low MoCA scores suggest the presence of mild cognitive impairment. MOT performance decreased with age, and there were statistically significant differences between the young and elderly group overall regarding MOT results. Within the elderly group, statistically significant differences were observed between the NM and LM group, with the NM group obtaining better MOT results than the LM group. As such, the MOT is able to differentiate WM performance of young and older participants (with/without cognitive decline). In each group, the MOT proved to be a satisfactory WM measure. Within the three groups, the obtained results showed adequate variability, and ceiling or floor effects were not detected.

These different studies suggest that the MOT might be seen as a general purpose instrument, useful for the assessment of WM in healthy and clinical populations, regardless of the age range.

II - Aims of the Present Work

The studies presented in this thesis were conducted to determine the appropriateness of using the MOT as a tool for assessing WM in adults across a broad age range, and to provide its potential users with normative data to ground such assessments. Accordingly, we sought to obtain construct-related evidence of validity for the MOT, as well as to determine its reliability (test-retest and internal consistency). As this instrument revealed to have adequate validity and reliability, we completed the collection and duly organized normative data for use with speakers of Continental Portuguese. More specifically, we sought: 1) to obtain convergent and discriminant validity evidence for the MOT, by correlating this instrument with the Daneman and Carpenter RST (Daneman & Carpenter, 1980) and the backward condition of the DS (Wechsler, 2008) (convergent evidence), as well as with measures probably less related to the construct (discriminant evidence), such as measures of inhibitory efficiency (derived from performance on the Stroop Neuropsychological Screening Test) (Stroop, 1935; Trennery, Crosson, DeBoe, & Leber, 1985; Castro, Cunha, & Martins, 2000), measures of processing speed (speeded color naming) and with the forward condition of the DS; 2) to evaluate the MOT's test-retest reliability by assessing 40 participants twice (the testing sessions being separated by a 12-14 week interval), and internal consistency by computing item-total correlations and Cronbach's alpha; 3) to establish norms for using the MOT with speakers of European Portuguese. Extensive analyses were performed to elucidate the influence of age, education and gender on MOT performance.

III - Method

Firstly, the participants that integrated each one of the studies will be characterized, as well as the procedures involved in data collection and the tasks that were administered. In the last part of this section a reference to the statistical analyses will be made.

1. Participants

1.1. Convergent and Discriminant Validity Studies

Aspects of MOT construct validity were evaluated by using data gathered from 81 participants. The sample consisted of 33 women (41-76 years, mean age=52.58) and 48 men (30-76 years, mean age=43.94), with overall 3-20 years of formal education (*M*=11.432; *SD*=4.39). Potentially healthy participants were approached by the researchers with the intention of including a broad range of adult ages and education levels. Participation was voluntary and confirmed by written consent. Inclusion criteria comprehended auditory and visual acuity (self-reported), which should be either normal or corrected-to-normal. Participants were also expected to be native speakers of European Portuguese and to have completed at least three years of formal schooling. Exclusion criteria consisted of a history of pre-existing psychiatric, neurological or cardiovascular disorders or of substance abuse or dependence, as well as of head injury or any other medical condition that might affect performance. Participants with ages equal or higher than 45 years were administered the Portuguese version of the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005; Freitas, Simões, Martins, Vilar, & Santana, 2010). Those scoring at least two standard deviations below their age group and education level were excluded from the study.

In order to assess discriminant validity, measures of inhibitory efficiency (derived from the Stroop Neuropsychological Screening Test) and processing speed (derived from a speeded color naming task) were collected. The forward condition of the DS, as a measure of short-term memory span, was used to complement the discriminant validity evidence. Besides the MOT, the Daneman and Carpenter RST and the backward condition of the DS were employed to evaluate WM. Data from these WM tasks was used to explore the MOT's convergent validity.

In order to keep the duration of the assessment session fairly short and acceptable to all participants, the DS was performed by the participants who did not have to complete the MoCA, i.e., those younger than 45 years. The DS was also completed by the persons that participated in the test-retest study. This resulted in a sample of 48 participants, comprising 18 women (41-50 years, mean age=44.94) and 30 men (30-50 years, mean age=38.93), with overall 4-20 years of formal education (*M*=12.896, *SD*=4.054).

1.2. Reliability Study

A sub-sample of the participants that took part in the convergent and discriminant validity studies was invited to also partake in a test-retest reliability study. The sample consisted of 40 individuals within the 30-50 age range, comprising[1] 18 women (41-50 years, mean age=44.94) and 22 men (30-50 years, mean age=39.45), with overall 4-20 years of formal education (*M*=13.10, *SD*=4.088). The second data gathering session took place within an interval of 12-14 weeks after the first.

The consistency of participant's responses across the MOT's items at one point in time was also evaluated, by calculating item-total correlations and Cronbach's alpha. To estimate internal consistency, alpha coefficients were obtained for each administration of the MOT (using the data of the participants that integrated the test-retest study). Alpha coefficients for the normative data sample were equally obtained.

1.3. Normative Data

MOT results from several studies were merged in order to obtain a normative sample. Besides the data from the aforementioned 81 participants, results from three other studies with 166 participants in the 18-80 age range who completed the MOT as part of the respective assessment protocols were equally included to enhance the normative sample regarding age-range representativeness. The participants from these studies were screened either with the Portuguese versions of the MoCA or Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975; Morgado, Rocha, Maruta, Guerreiro, & Martins, 2009)[2]. Using the adequate age/education cut-off points for each instrument, this resulted in the

[1] A chi-square goodness-of-fit test showed that there was no statistically significant difference between the male/female ratio [$\chi^2(1)=0.40$, $p=0.527$].

[2] The exclusion criteria described in the previous section were also applied.

exclusion of one participant (MMSE score fell below the cut-off point). This resulted in a sample of 247 participants, 140 women (18-78 years, mean age=42.078) and 107 men (20-80 years, mean age=52.556), with 3-24 years of formal education (*M*=10.967, *SD*=4.313).

2. Procedure

2.1. Convergent and Discriminant Validity Studies

The relation between performance on the MOT and performance on other tasks more frequently used to assess WM capacity was examined with respect to the Daneman and Carpenter RST (Daneman & Carpenter, 1980) and the DSB (Wechsler, 2008). Discriminant validity evidence was provided by measures of inhibitory efficiency using the Stroop Neuropsychological Screening Test (SNST, Stroop, 1935; Trennery et al., 1985; Castro et al., 2000) and by measures of processing speed, employing a speeded color naming task (which we developed on basis of the SNST materials). The DSF (Wechsler, 2008) also entered the discriminant validity study. Hence, the neuropsychological protocol used in this study included the MOT, the SNST (comprising two additional subtasks of black-ink color reading and speeded color naming), a Portuguese rendition of the Daneman and Carpenter RST, and (for participants younger than 45 years and for the test-retest study participants) the DS. So as to control for order effects, the order of presentation of the MOT and the RST was counterbalanced across participants.

The MOT requires participants to put into calendar sequence increasingly long sets of months that are verbally presented out of calendar order. The task contains five different span levels (varying between 2 and 6 months), and each span level contains four trials. The task starts with two-item trials, and ends with six-item trials. The totality of the trails was presented to the testees, and the participants' final score was composed by the number of correct answers. The raw final score was transformed into a proportion-correct score (raw MOT score ÷ 20). Concerning the SNST[3], participants are first presented with two practice sheets in order to ensure that they are able to correctly name the four colors used in the test, that the font size allows perfect reading and that the overall task instructions are understood. After the pre-test, the participant is given a sheet containing 112 stimuli arranged in four columns. The stimuli consist of four Portuguese color names (*rosa, verde, azul, cinza*), and each word is printed in incongruous ink colors. Firstly, the participant is asked to read the words down the columns as quickly as possible. In the second condition, and using the same sheet, the individual is required to name the color of the ink in which the different words are printed, also as rapidly as possible. The time that the participant needs to complete each of the tasks is measured with a stopwatch. The standard SNST was complemented with two additional tasks. In the first complementary subtest, the participant is shown a new

[3] Please see Results section 1.2. for a detailed description of the obtainment of the inhibitory and processing speed measures and results.

sheet on which the original SNST stimuli are printed in black ink. The individual has to read these words as quickly as possible. The other subtest is a speeded color naming task, in which the printed words are substituted by colored bars, maintaining the original colors and number of stimuli. The color of each bar has to be named as quickly as possible. Again, both conditions are timed with a stopwatch. The speeded color naming task allows us to obtain a measure of processing speed. In the Daneman and Carpenter RST, participants are required to read aloud sentences that appear on a computer screen. The sentences remain on the screen during five seconds. Participants are instructed to remember the final word of each sentence for later recall. After the presentation of a set of sentences, the participant has fifteen seconds to recall the target words in the same order in which they were displayed. Set size is gradually increased, creating five span levels of two, three, four, five, or six sentences (each one containing 13-16 words) that have to be read before engaging in the memory task. Each span level contains three groups of sentences, paired with three corresponding recall tasks. The participants' final score was the number of correct answers. The raw final score was transformed into a proportion-correct score (raw RST score ÷ 15). The session ended with either the DS, for participants younger than 45 years, or with the MoCA for the older participants. The DS included two conditions: DSF and DSB. In the DSF part of the test, the participant is read a list of numbers (at a rate of one per second), and is expected to repeat the number string in the exact same order. The lists gradually increase in length and, consequently, in difficulty, varying between two and nine numbers. DSF consists of eight pairs of number sequences. In the backward condition, the participant is equally cited a set of digits, this time being required to repeat the numbers in reverse order. The lists also increase in length and difficulty, containing a maximum of eight numbers. DSB comprises seven pairs of numbers. The participants' final score comprised the number of correct answers. The raw final score was transformed into a proportion-correct score (raw DSF score ÷ 16, raw DSB score ÷ 14).

2.2. Test-retest Reliability

Forty participants who took part in the convergent and discriminant validity were once again assessed, 12-14 weeks after the first testing session. Special care was taken in order to insure that administration conditions were as similar as possible to those of the first session, including the starting time and the order of presentation of the two WM tasks. Participants who had experienced changes in terms of physical or psychological health (e.g., medical surgeries, prescription of antidepressants or sleeping pills), or changes concerning their personal life (e.g., divorce, loss of a loved one) were excluded from this study. Data from participants whose second evaluation took place in an environment that substantially differed from the setting of the first testing session (e.g., presence of distractors) were equally discarded. The session started with an update of the information previously gathered with the semi-structured interview. Before the WM tasks, the SNST and its supplementary tasks were applied anew, and the session ended with the DS.

2.3. Normative Data

Normative data for the MOT resulted from merging the data provided by the different studies described above. In all of these studies the MOT was administered in the context of a short battery of neuropsychological tests. The position of the MOT within the neuropsychological testing sessions varied within and across studies, sampling a wide range of administration positions. All of the batteries that were administered to participants over 40 years old included either the MoCA or the MMSE, which we used to implement our exclusion criteria regarding possible mild dementia.

3. Statistical Analyses

All the analyses were conducted using IBM SPSS statistics 20.0 (SPSS Inc., Chicago, IL), and the level of statistical significance was set at p=0.05.

3.1. Convergent Validity Evidence Study

In order to obtain convergent validity evidence for the MOT, Pearson product-moment correlations were calculated between the MOT scores, RST scores and DSB scores. Proportions of correct responses were used in all computations. Pearson product-moment correlations between RST scores and the DSB were equally obtained. These data were used for comparative purposes regarding the MOT's correlation coefficients. Cohen's (1988) conventions[4] were used to interpret effect size. Proportion-correct scores are binomially distributed, being therefore inappropriate for tests of statistical significance. So as to approximate a Gaussian distribution, the arcsin square root transformation was applied to the proportion-correct scores (Sokal & Rohlf, 1995).

In regard to univariate outliers, boxplots were employed to identify deviant data points. Extreme outliers[5] were winsorized, whereas moderate outliers were treated like any other data point. Nevertheless, in case the moderate outliers affected the strength and/or significance of the correlation between the variables, winsorizing was also performed. Normality analyses were equally implemented. A distribution was considered normal when the Shapiro-Wilk test was not significant and/or the absolute value of the skewness was not equal to or more than twice the standard error. If any of these criteria was not met, a non-parametric test was additionally carried out (Spearman's rho).

3.2. Discriminant Validity Evidence Study

So as to attain discriminant validity evidence for the MOT, we computed Pearson product-moment correlations between the proportion-

[4] Cohen (1988) provides the following rules of thumb: $r \geq (-)0.10$ – small effect size; $r \geq (-)0.30$ – medium effect size; $r \geq (-)0.50$ – large effect size.
[5] Extreme outliers are data points that lie 3 *SD* above/beneath the mean.

correct scores of the MOT, proportion-correct scores of the DSF task and measures of inhibitory efficiency derived from the SNST. Measures of inhibitory efficiency are expressed as the difference between incongruent and neutral trials (MacLeod, 1991). In order to include all possible measures of inhibition not only Stroop interference scores were estimated (i.e., ink-color naming impeded by the incongruent colored word), but also scores that could potentially reflect Stroop's complementary interference effect (i.e., colored-word reading impeded by the incongruent ink color). Furthermore, interference scores were obtained not only considering the participants' performance times, but also taking into account the proportion-correct scores obtained in each condition. Discriminant validity evidence also included Pearson product-moment correlations between proportion-correct scores of the MOT and measures of processing speed (derived from a speeded color naming task). Once again, Pearson product-moment correlations between proportion-correct scores of the RST and the aforementioned measures were obtained as well (for comparative purposes). Cohen's (1988) conventions were used to interpret effect size (please see footnote 4).

The arcsin root square transformation was applied to the proportion-correct scores, and the procedures from the previous study regarding normality and univariate outliers were adopted.

3.3. Reliability Study

The MOT's test-retest reliability was estimated by calculating Pearson product-moment correlations between MOT proportion-correct scores obtained during two distinct testing sessions (separated by a 12-14 week interval). The RST's retest coefficient was estimated as well (for comparative purposes).

The MOT's internal consistency was assessed by item-total correlations and Cronbach's alpha coefficient (using the raw total scores). Alpha Coefficients were obtained for each administration of the task, for both the retest and normative samples. So as to interpret the magnitude of the alpha coefficients, the criteria[6] described by Pestana and Gageiro (2008) were taken into consideration.

3.4. Influence of Demographic Characteristics on MOT Performance

So as to compute and present adequate MOT norms for use in clinical settings, an interruption criterion for the MOT was introduced to obtain each participant's final score. We recommend interruption of the task after the testee fails four consecutive trials on the same level[7]. Therefore, the analyses

[6] Internal consistency is considered excellent when $\alpha>0.9$; good when $\alpha>0.8$ and $\alpha<0.9$; reasonable when $\alpha>0.7$ and $\alpha<0.8$; weak when $\alpha>0.6$ and $\alpha<0.7$; and unacceptable when $\alpha<0.6$.

[7] We consider Almor and colleagues' (1999, 2001) interruption criterion (test administration is stopped after the participant fails two consecutive items on the same level) somewhat restrictive, as the majority of our participants produced correct answers after two successive errors. A large portion (41%) of our normative sample (N=247) was able to give accurate answers (ranging between 1 and 5) even after commiting two sequential errors on the same level. As such, the interruption criterion proposed by Almor and colleagues might bring a loss

that intended to assess the contribution of demographic variables on MOT performance and the normative data that we present for use with European Portuguese speakers (section 3.5.) took the proposed interruption criterion into account.

A hierarchical multiple regression analysis was used to assess the contribution of demographic variables (age, gender, years of formal education) to performance on the MOT. The arcsin root square transformed proportion-correct score of the participants until erring four consecutive trials on a same level was entered as dependent variable, and age, gender and years of formal education were entered as predictors. Linear and quadratic terms were included for age and years of formal education (setting a maximum of 17 years for the latter variable so as to eliminate outliers). Age and years of formal education were centered in an attempt to avoid multicollinearity (Marquardt, 1980). Gender was coded as female=0 and male=1. Thus, the MOT proportion-correct scores were regressed on age, age squared, gender, years of formal education and years of formal education squared. The hierarchical regression analysis included five levels. The first block included the age variable. Age squared was entered in the second model and the gender variable was included in the third model. The participant's level of education was entered in the fourth step and its quadratic term (years of formal education squared) was included in the last (fifth) block. Non-significant predictors were left out in a subsequent, non-hierarchical regression analysis. Influential observations were identified by calculating Cook's distances and Leverage values. Participants with Cook's distances equal to or superior to $4/n$ and/or Leverage values equal to or superior to $2(k+1)/n$ were excluded from the subsequent analysis. Normality, multicollinearity and homoscedasticity were inspected. A distribution was considered normal when the Shapiro-Wilk test was not significant and/or the absolute value of the skewness was not equal to or more than twice the standard error. A final, non-hierachical linear multiple regression analysis was performed with the untransformed proportion-correct MOT scores (using our interruption criterion) as dependent variable and age, gender and years of formal education as predictors, in accordance with the significance

---

of information regarding WM performance. Nevertheless, submitting testees to the totality of the task (regardless of the stage where the examinee fails to produce correct responses) might also not be beneficial, especially in the case of older/cognitively impaired patients. In our sample, we have observed that, after commiting 4 errors on the same level, participants rarely produce correct responses. Only 3% of our sample was able to deliver correct answers (ranging between 1 and 2). Hence, we suggest task interruption after the participant errs all four trials of a certain level. The strong and positive correlation between the MOT's total proportion-correct scores and the proportion-correct scores computed after truncation using our interruption criterion [$r(245)=0.968$, $p<0.001$; $rs(245)=0.996$, $p<0.001$] illustrates that the information that is lost after task interruption is negligible. Furthermore, the interruption criterion does not compromise the temporal stability of the MOT, as an high and positive correlation has been observed between the proportion-correct scores based upon the interruption criterion, computed for the test-retest sessions [$r(38)=0.918$, $p<0.001$; $rs(38)=0.887$, $p<0.001$]. These proportion-correct scores are also strongly related to other measures of WM. A large and positive correlation was attained with the RST [$r(79)=0.596$, $p<0.01$; $rs(79)=0.567$, $p<0.001$] and DSB [$r(46)=0.505$, $p<0.01$].

results obtained in the first analysis. The variance inflation factors (VIFs) were used as indicators of multicollinearity and should not be superior to 10 (Belsley, Kuh, & Welsch, 1980). Homoscedasticity was evaluated through scatterplots. Multicollinearity was absent and the assumption of homoscedasticity was met. The regression coefficients were used to compute regression-based norms, as described in the following section (section 3.5.).

So as to further explore the impact of age on MOT scores and to examine the possible differences in MOT performance between the various age groups, a two-way between-groups analysis of covariance was conducted. The statistical analysis, results and discussion of this accessory study can be consulted in Appendix III.

3.5. Regression-based norms for the MOT

In order to provide the potential users of the MOT with a means to control for all the demographic variables that influence MOT performance, regression-based norms were obtained. A regression-based algorithm was assembled, giving the user the opportunity to standardize the obtained MOT scores, taking into account the influence of age, gender and years of formal education. The algorithm generates a Z-score expressing how the difference between a testee's actual score and the score predicted on the basis of his/her gender, age and years of formal education compares with this same difference in the normative sample. The algorithm includes the regression coefficients calculated for each relevant demographic variable, and, to implement the standardization of the testee's residual, the standard deviation of the residuals in the normative sample. Individual scores can be standardized by following these steps (Van Breukelen & Vlaeyen, 2005; Van der Elst, Dekker, Hurks, & Jolles, 2012): 1) the testee's predicted MOT score is calculated by means of the final regression model[8]; 2) the residuals of the score are computed (residual = observed value – predicted value); 3) The participant's residual is standardized (the residual is divided by the standard deviation of the residuals of the normative sample).

IV – Results

Table 1. Descriptive statistics for measures of memory, inhibition and processing speed, using untransformed proportion-correct scores and measures of speed, expressed as inverse duration in milliseconds (ms-1) or syllable rate (syllables per second).

	M (SD)	*N*
Memory tasks		
1. MOT	0.709 (0.138)	81
2. RST	0.348 (0.128)	81
3. DSF	0.577 (0.127)	48
4. DSB	0.467 (0.161)	48

[8] Expected MOT score = $B_0 + B_1X_{1i} + \ldots + B_kX_{ki}$ (with B_0 being the intercept, B_k the regression weights for the demographic variables and X_{ki} the values of the demographic variables for a certain participant) (Van der Elst et al., 2012).

Inhibitory efficiency measures[9]		
5. RCWpc-RBWpc	-0.257 (2.010)	81
6. RCWs-RBWs (ms^{-1}**)**	-0.001 (0.002)	81
7. NCWpc-NCSpc	-3.053 (5.241)	81
8. NCWs-NCSs (ms^{-1}**)**	-0.004 (0.001)	81
Measures of processing speed[9]		
9. NCSs (ms^{-1}**)**	0.012 (0.002)	81
10. NCSsr (syll. per sec)	2.693 (0.501)	81

Notes. M=mean, *SD*=standard deviation, *N*=sample size.

1.1. Convergent Validity Study

So as to obtain convergent validity evidence for the MOT as a measure of working memory, Pearson-product moment correlations were calculated between MOT and RST scores, and between MOT and DSB scores. Correlations were also computed between the RST and DSB scores. Proportions of correct responses were used in all computations, and the arcsin square root transformation was applied. This study's sample included 81 participants with ages ranging between 30 and 76 years (*M*=47.46, *SD*=12.339). The sample that completed the DSB consisted of 48 participants with ages varying between 30 and 50 years (*M*=41.19, *SD*=6.149).

A large, positive correlation was observed between the MOT and RST [$r(79)=0.589$, $p<0.01$][10]. A strong and positive relation between the MOT and the DSB was also observed [$r(46)=0.513$, $p<0.01$][11]. Lastly, Pearson product-moment correlations were computed between the RST and DSB. A medium and positive relation between the variables was found [$r(46)=0.402$, $p<0.01$].

The correlation matrix can be found in Appendix I.

1.2. Discriminant Validity Study

In order to establish discriminant validity evidence for the MOT as a measure of working memory, Pearson product-moment correlations were estimated between MOT and DSF proportion-correct scores; MOT proportion-correct scores and measures of inhibitory efficiency, and MOT proportion-correct scores and measures of processing speed. As previously mentioned in the statistical analysis section, measures of inhibitory efficiency are derived from the SNST and are expressed as the difference between incongruent and neutral trials. The following measures of inhibitory efficiency were thus obtained: 1) RCWpc-RBWpc: difference in proportion-correct scores for reading color names presented as (incongruent) colored

[9] Please see the Acronyms List (or Results Section 1.2.) for acronym meanings.

[10] Due to the violation of the normality assumption underlying statistical significance testing (although not Person's r magnitude and its interpretation) Spearman's rho was also computed as cautionary measure against type I errors. A significant correlation was equally obtained [$r_s(79)=0.559$, $p<0.01$].

[11] As the assumption of normality was once again not met, Spearman's rho was calculated. The correlation between the variables was significant as well [$r_s(46)=0.544$, $p<0.01$].

words (RCWpc) vs. black words (RBWpc); 2) RCWs-RBWs: difference in speed[12] while reading color names presented as (incongruent) colored words (RCWs) vs. black words (RBWs); 3) NCWpc-NCSpc: difference in proportion-correct scores for naming colors presented as colored words (incongruous hue/color names) (NCWpc) vs. colored squares (NCSs); 4) NCWs-NCSs: difference in speed for naming colors presented as colored words (NCWs) representing incongruous color names vs. colored squares (NCSs). Two measures of processing speed were derived from a speeded color naming task: 1) NCSs: naming speed of colored squares; 2) NCSsr: syllable rate[13] (syllables per second) while naming colors presented as colored squares.

The proportion-correct scores were transformed using the arcsin square root transformation. Pearson product-moment correlations were also calculated between the above mentioned measures and the RST proportion-correct scores. The sample consisted of 81 participants, with a 30-76 age range (M=47.46, SD=12.339). The sample that completed the DSF task contained 48 participants, with ages varying between 30 and 50 years (M=41.19, SD=6.149).

The MOT and the DSF were positively and significantly correlated [$r(46)=0.617$, $p<0.01$][14]. A moderate and positive correlation was observed between the RST and DSF [$r(46)=0.395$, $p<0.01$].

Regarding the relations between the MOT and measures of inhibitory efficiency, the observed correlations were small or non-significant. The correlation between the MOT and the first measure of inhibitory efficiency (RCWpc-RBWpc) was non-significant [$r(79)=-0.032$, $p=0.775$][15]. The correlation between the MOT and RCWs-RBWs was only marginally significant [$r(79)=-0.210$, $p=0.060$][16]. A small and positive relationship was observed between the MOT and NCWpc-NCSpc [$r(79)=0.254$, $p<0.05$]. The relation between the MOT and the last variable of inhibitory efficiency (NCWs-NCSs) was non-significant [$r(79)=-0.021$, $p=0.852$][17].

Correlations between the RST and measures of inhibitory efficiency were also computed, with the coefficients' magnitudes varying between small and medium values, or being non-significant. The RST and RCWpc-RBWpc were not significantly correlated [$r(79)=0.081$, $p=0.474$][18]. The correlation between the RST and RCWs-RBWs was small and negative [$r(79)=-0.259$, $p<0.05$]. A positive and moderate relationship was observed between the RST and NCWpc-NCSpc [$r(79)=0.305$, $p<0.01$]. The relation

---

[12] Speed measures were obtained the following way: 1 ÷ performance time.
[13] The syllable rate was calculated the following way: (number of syllables per color name × number of correct responses) ÷ performance time.
[14] Spearman's rho was equally calculated, as the assumption of normality was violated. The correlation remained significant [$r_s(46)=0.600$, $p<0.01$].
[15] The assumption of normality was not met, so Spearman's rho was also estimated. The variables remained non-significant [$r_s(79)=0.016$, $p=0.891$].
[16] A violation of the assumption of normality occured. Therefore, Spearman's rho was carried out. The variables were not significantly correlated [$r_s(79)=-0.184$, $p=0.101$].
[17] Spearman's rho was equally obtained, as a violation of the normality assumption occurred. No significant correlation was observed [$r_s(79)=-0.040$, $p=0.724$].
[18] As the normality criteria were not met, Spearman's rho was calculated. The two variables were not significantly correlated [$r_s(79)=0.045$, $p=0.678$].

between the RST and the last measure of inhibitory efficiency (NCWs-NCSs) was non-significant [$r(79)=0.099$, $p=0.379$].

Correlations between the MOT and measures of processing speed were also calculated. A medium and positive correlation was uncovered between the MOT and NCSs [$r(79)=0.353$, $p<0.01$][19], as well as between the MOT and NCSsr [$r(79)=0.342$, $p<0.01$].

Lastly, the relation between the RST and measures of processing speed were examined. The RST was moderately and positively correlated with NCSs [$r(79)=0.429$, $p<0.01$][20], and also with NCSsr [$r(79)=0.426$, $p<0.01$][21].

The correlation matrix can be found in Appendix I.

1.3. Reliability

In order to establish test-retest reliability for the MOT, Pearson product-moment correlations were computed between the transformed proportion-correct scores of the MOT obtained on two different testing sessions (separated by 12-14 weeks). The test-retest sample comprised 40 participants, with ages varying between 30 and 50 years (M=41.93, SD=6.220). The two variables were strongly and positively correlated [$r(38)=0.925$, $p<0.01$]. Pearson-product moment correlations were also calculated between the transformed proportion-correct scores of the RST, also obtained on the two distinct testing sessions. A strong and positive relation was equally observed [$r(38)=0.640$, $p<0.01$].

Table 2. Descriptive statistics for and test-retest reliability of the MOT and RST (Proportion-correct scores were used).

	M (SD) **Time 1**	***M (SD)*** **Time 2**	**Time 1 - Time 2 correlation**
MOT	0.748 (0.137)	0.764 (0.122)	0.925**
RST	0.378 (0.146)	0.402 (0.157)	0.640**

Notes. N=40. Time 1: scores obtained during the first testing session, Time 2: scores obtained during the second testing session. M=mean, SD=standard deviation. ** p<0.01

In order to examine the internal consistency of the MOT, Cronbach's alpha was computed using the test-retest sample, so as to study the internal consistency of the scale across two different points in time. The alpha coefficient using the entire normative sample was also calculated. The participants' raw MOT scores were used.

The internal consistency results of the MOT using the data from the test-retest sample are displayed in table 3.

[19] Spearman's rho was also estimated, due to the violation of the normality assumption. The variables were moderately and positively correlated [$r_s(79)=0.303$, $p<0.01$].
[20] The assumption of normality was violated, so Spearman's rho was additionally computed. A moderate and positive correlation was equally unveiled [$r_s(79)=0.353$, $p<0.01$].
[21] Due to a violation of the normality assumption, Spearman's rho was calculated. The variables were moderately and positively correlated [$r_s(79)=0.353$, $p<0.01$].

Table 3. Item-total correlations and alpha values for each administration of the MOT using the test-retest sample.

	Correlations of item with total	
Set size	**Time 1**	**Time 2**
Three	-0.038	-
Four	0.531	0.486
Five	0.727	0.474
Six	0.473	0.375
Alpha	0.648	0.629
Alpha adjusted	0.739	-

Note. Alpha adjusted is with set size three omitted. Time 1: first administration of the MOT, time 2: second administration of the MOT.

Concerning the internal consistency results for the first administration, set size two was automatically excluded as the variable had zero variance. When estimating the alpha coefficient with set sizes three, four, five and six, an alpha value of 0.648 was obtained. This value might be considered weak. However, when set size three was omitted an alpha of 0.739 was attained, assuming a reasonable value. When estimating Cronbach's alpha for the second (retest) administration, a weak result was achieved (α=0.629). Set sizes two and three were automatically excluded from the anaylsis due to a lack of variance.

The internal consistency[22] for the MOT using the normative sample was also assessed. After the exclusion of one participant (due to a low MMSE score), the normative sample included 247 participants, with an age range of 18-80 years (*M*=46.664, *SD*=19.073). Results are summarized in table 4.

Table 4. Item-total correlations and alpha values for the MOT using the normative sample.

Set size	**Correlations of items with total**
Two	0.141
Three	0.351
Four	0.597
Five	0.675
Six	0.551
Alpha	0.686
Alpha adjusted	0.727

Note. Alpha adjusted is with size two omitted.

When set size two was included, a weak alpha coefficient was obtained (α=0.686). Nevertheless, when set size two was excluded from the analysis, the alpha coefficient reached an acceptable value (α=0.727).

1.4. Influence of Demographic Characteristics on MOT Performance

The contribution of demographic variables (age, gender, years of formal education) to MOT performance was examined employing a hierarchical multiple linear regression analysis. The hierarchical multiple

[22] The internal consistency of the MOT when using the interruption criterion was equally estimated. When set size two was included in the analysis, an alpha coefficient of 0.694 was achieved. After excluding this first item, an alpha coefficient of 0.736 was observed.

regression models can be consulted in Appendix II. Subsequently, a final non-hierarchical linear multiple regression analysis was performed, with the untransformed proportion-correct MOT scores (applying our interruption criterion) as dependent variable and age, gender and years of formal education as predictors (in accordance with the significance results obtained in the hierarchical regression analyses) (table 5). Data from 15 participants was excluded from the analyses, as the Cook's distances', Leverage values' and MMSE cut-off points were exceeded. This resulted in a sample of 232 participants, with ages varying between 18-80 years (*M*=45.522, *SD*=18.947).

The results indicate that age [β=-0.335, $p<0.001$], gender [β=0.127, $p<0.05$] and years of formal education [β=0.398, $p<0.001$] each made unique contributions to MOT performance.

The model as a whole is significant [$F(3, 228)=52.681$, $p<0.001$], explaining 40.9% of the variance.

Table 5. Non-hierarchical linear multiple regression analysis for variables (age, gender, years of formal education) contributing to MOT performance.

	B	***SE***	***β***
Constant	0.592	0.046	
Age	-0.003	0.001	-0.335***
Gender	0.040	0.018	0.127*
Education	0.016	0.003	0.398***

Note. R^2=0.409. R^2=R squared, *B*=unstandardized regression coefficient, *SE*=standard error, *β*=standardized regression coefficient. *p<0.05, ***p<0.001

1.5. Regression-based Norms for the MOT

So as to control for the influence of demographic variables (age, gender and years of formal education) on MOT performance, regression-based norms for speakers of European Portuguese were established. The demographic structure of the normative sample's and the corresponding stratified proportion-correct scores (untransformed and truncated using our interruption criterion), are displayed in table 6.

Table 6. Demographic structure of the normative sample and stratified descriptive statistics for the untransformed MOT proportion-correct scores (considering the interruption criterion).

	N	**%**	***M(SD)***	**MIN-MAX**
Gender				
Male	104	44.83	0.666(0.161)	0.30 – 1.00
Female	128	55.17	0.652(0.158)	0.30 – 1.00
Age				
18-39	83	35.78	0.741(0.115)	0.50 – 1.00
40-59	77	33.19	0.687(0.149)	0.40 – 1.00
60-69	37	15.95	0.557(0.143)	0.30 – 0.85
70-80	35	15.09	0.509(0.125)	0.30 – 0.80
Education				
≤4	33	14.22	0.481(0.116)	0.30 – 0.75

5-9	50	21.55	0.598(0.139)	0.35 – 1.00
10-12	109	46.98	0.696(0.127)	0.35 – 1.00
12+	40	17.24	0.778(0.141)	0.45 – 1.00

Notes. Influential data point were excluded (based on Cook's distances and Leverage values), *n*=sample size, *M*=mean, *SD*=standard deviation, MIN=minimum, MAX=maximum.

A regression-based algorithm was obtained so that MOT scores can be adjusted for gender, age, and years of formal education. This equation enables the user to transform proportion-correct MOT scores into standardized Z scores. The following equation was obtained[23]:

Z MOT = (proportion-correct MOT score – 0.592 + (0.003 × age) – (0.040 × gender) – (0.016 × years of formal education)) ÷ 0.12216

Notes: proportion-correct score = obtained MOT score ÷ 20. Gender is coded as female = 0 and male = 1. The regression weights can be consulted in table 5.

V – Discussion

The present study intended to obtain convergent validity evidence, discriminant validity evidence and reliability data for the MOT. We equally aimed to establish regression-based normative data for speakers of European Portuguese.

Concerning the results of the convergent validity studies, large and positive correlations were observed between the MOT and the two other WM measures (RST and DSB). The magnitude of the correlation coefficients are congruent with our expectations, as positive and significant correlations were foreseen. The DSB and the RST were only moderately correlated. The results suggest that the aforementioned WM tasks measure a similar construct, yet are not identical. Several authors (e.g., Conway et al., 2005; Waters & Caplan, 2003) claim that WM tasks are not process pure, and "(…) no single task is a perfect measure of the construct it ostensibly represents" (Conway et al., 2005, p. 780). For instance, the RST is a widely used WM measure, but it also taps verbal ability (Daneman & Carpenter, 1980). The DSB task might also imply a visuo-spatial component, as participants may attempt to visualize the numbers whilst executing the task (Reynolds, 1997). In fact, the overall pattern of correlations among the three

[23] Let us consider the example of a female participant with 6 years of formal education and 70 years of age that has obtained a raw MOT score of 13 points. This raw score can be transformed into a proportion-correct score: 13/20=0.65. In order to obtain a standardized score, the user is required to complete the equation with the examinee's proportion-correct score and demographic values. Using the data of the aforementioned participant, the following equation is obtained: Z MOT = (0.65 – 0.592 + (0.003 × 70) – (0.040 × 0) – (0.016 × 6)) ÷ 0.12216
The result is a standardized score of approximately 1.41 (Z values must be recorded to two decimal places). We can conclude that the participant's result lies between 1 and 2 SD above the mean (Z scores have a distribution with a mean of 0 and a SD of 1). After consulting a Z distribution table, it is possible to note that nearly 92% of 70 years old women in the population, with 6 years of formal education, have equal or lower MOT scores.

different WM tasks that we have obtained (ranging between 0.402 and 0.589) is consistent with those attained in a wide array of other studies. Conway, Cowan, Bunting, Therriault and Minkoff (2002) have equally observed moderate correlations (0.49 – 0.52) between several WM tasks. Engle, Tuholski, Laughlin and Conway (1999) have attained similar results (with correlations varying between 0.32 and 0.51). In accordance with the observed results, Waters and Caplan (2003) pointed out that performance on WM tasks is mostly only moderately correlated.

The lowest correlation was observed between the RST and the DSB (r=0.402). Various other studies have attained comparable outcomes. For instance, Park and colleagues (2002) found a 0.50 correlation between the RST and the DSB. By means of an exploratory factor analysis, Oberauer and coworkers (2000) suggested that RST and DSB loaded on the same factor (representing verbal WM). Nevertheless, this moderate relation between RST and DSB might be explained by the fact that the RST implies heavier processing demands, encompassing an overt sentence processing component. As such, this task demands a wide range of language comprehension abilities (MacDonald et al., 2001). Despite the fact that DSB involves mental manipulation (by requiring the digits to be arranged in reverse order), its processing requirements are lighter than those recruited by the RST, as it does not involve the reading of sentences[24]. According to Almor and colleagues (2001), the MOT relies on language comprehension processes as well, since "(…) the participant must process a speech signal (the list of months), maintain a representation of the input in mind, convert it to a semantic representation (i.e., establish that the phonological string "June" means the month of June), and use this representation to reorder the input and formulate the response" (p. 46). Even though the DSB task equally implies language comprehension processes, it does so to a lesser extent. In contrast with the MOT, the semantic information (the meaning of the individual digits) does not play a key role in the DSB task, as the testee is not required to reorder the digits, but simply has to repeat them in an inverted order (MacDonald et al., 2001). Yet, the MOT and the DSB present some similarities, as they are not based upon a complex dual task paradigm and involve fairly simple instructions (unlike the RST). These resemblances might be responsible for the strong correlation between the MOT and the DSB. The highest correlation was attained between the MOT and the RST (0.589). These two tasks share some similarities that might explain this correlation coefficient. As already discussed, both tasks rely quite strongly on language comprehension processes, although for different reasons. The RST triggers these processes due to its reading component, whereas semantic information is essential to MOT performance. In spite of these differences, it is likely that both the MOT and RST assess, besides WM, verbal abilities. Also, the MOT and the RST show resemblances in terms of

[24] As reading is an automatized process, comprehension processes are likely to be triggered during the act (whether one intends it or not). Comprehension is a complex process and depends upon various skills, its discussion being beyond the scope of this thesis.

administration procedures and task design (Almor et al., 1999, 2001). Both tasks contain five different span levels (varying between 2 and 6 to-be-remembered words or months), and each span levels comprises four trials. Even though the tasks differ in terms of their processing demands (i.e., reading of sentences or ordering of a list of months), they both entail similar storage demands.

Everything considered, the results that we have obtained suggest that MOT performance relates well to that in other tasks that are thought to measure the same construct (WM).

In order to attain discriminant validity evidence for the MOT, we have interpreted the correlations between this WM task and measures of STM, inhibitory efficiency and processing speed. Correlations between the RST and these measures were also estimated. The MOT and the DSF appeared to be strongly correlated, whilst the RST and the DSF were only moderately related. Positive correlations between the WM tasks and STM task were anticipated, as both types of task measure the same basic processes to some extent (i.e., the storage component) (Dobbs & Rule, 1989). Conway and colleagues (2005) state that confirmatory factor analysis and structural equation modeling techniques have demonstrated that STM and WM tasks are highly correlated, albeit not strongly enough to suggest that WM and STM measure the same construct. Conway and colleagues (2002) found that, through structural equation modeling, STM and WM loaded on different factors. The STM memory tasks were positively correlated with the WM tasks (with correlations varying between 0.28 and 0.54). Engle and collaborators (1999) have reached similar conclusions. Confirmatory factor analyses showed that STM and WM are two different but highly related constructs (r=0.68). The correlation that we have obtained between the RST and the DSF is congruent with other authors' results. For instance, Barreyro, Burin, and Duarte (2009) reported a positive and moderate correlation between RST and DSF (r=0.416), as well as Park and colleagues (2002) (r=0.48). The high correlation between the MOT and the DSF that we attained might be explained by the fact that the MOT places great demands on storage capacity, especially during the first two trials. The MOT's initial items are fairly straightforward, as they comprise relatively few stimuli, and the participant does not have to greatly rely on his or her processing abilities in order to generate a correct response. However, according to Conway et al. (2005), "The early presence of difficult items may discourage some subjects, particularly those who are less able, such as children, the elderly, or patients" (p. 773). As such, the presence of these easier initial trials might be crucial in maintaining elder/cognitively impaired examinees motivated and willing to be assessed. These initial items might also serve another function, possibly being important in the participants' "warm-up", accustoming them to the task and preparing them for the subsequent (more complex) items. This might be particularly relevant in the assessment of older persons, as they generally possess a limited experience within a test-taking environment.

The discriminant validity study also entailed the computation of correlations between the MOT (and RST) and measures of inhibition and processing speed that were derived from the SNST. As already stated in the theoretical framework section, processing speed and inhibitory mechanisms play an important role in WM (and in other cognitive processes as well), and are thought to mediate age-related WM changes. Despite the relevance of these processes to WM performance, various studies have indicated that WM, inhibition and processing speed are separate (though related) constructs. McAuley and White (2011) carried out a latent variable analysis and found that (in a sample comprising children and young adults) processing speed, inhibition and WM were separable but related constructs. WM and processing speed were significantly correlated (r=0.57), as well as WM and inhibition (r=0.56).

As a consequence of the important role that inhibition and processing speed play in WM performance, significant correlations (yet inferior to those attained in the convergent validity study) were expected between the WM memory tasks and some of the measures of inhibitory efficiency and processing speed. In order to include all of the possible measures of inhibition that could be derived from the SNST materials, not only Stroop interference scores were estimated, but also scores that could potentially reflect Stroop's complementary interference effect (i.e., colored-word reading impeded by the incongruent ink color). Even though very few studies have reported a stable complementary interference effect, its existence is widely acknowledged (MacLeod, 1991). A speeded color naming task allowed us to obtain measures of processing speed. In addition to the naming speed for the colored squares, the syllable rate was also calculated (Olkkonen, 2013). Considering the measures of inhibitory efficiency, the correlations that we have obtained were mainly weak to moderate. The MOT was not significantly related to the measures reflecting Stroop's complementary interference effect (i.e., RCWpc-RBWpc and RCWs-RBWs)[25]. The RST was negatively and significantly correlated to the RCWs-RBWs measure (a small coefficient was obtained)[26].

The MOT revealed to be weakly but significantly related to the NCWs-NCSs[25] measure, and a medium and positive correlation between the RST and the same measure was observed. This moderate correlation between the RST and the NCWs-NCSs measure might be due to the fact that the RST is based on a complex dual-task design, placing quite high demands

---

[25] Please see the Acronyms List for the meaning of the variables.

[26] This result is not relevant for the purpose of the thesis and will, therefore, be discussed only briefly. The negative relation between these variables may express the fact that the slower the reading of the colored words (in comparison with the words printed in black ink), the better the RST scores. When reading words printed in incongruous colors, some participants may recruit greater strategic control over the reading process, lowering its normal automaticity level. This would allow intrusions of the ink color in word reading to be prevented. This propensity to adapt to the present task demands by shifting from automatic to more controlled processing might also be beneficial for RST performance. Even though this task implies the reading of sentences, the storage of the final-sentence words is the true purpose. Greater strategic control over the reading process (e.g., attempting to restrain the depth of processing of the irrelevant words and of the full sentence) may therefore benefit performance on RST.

on the participants. In these conditions, intrusions are more likely to occur, with participants recalling more non-target words (Robert, Borella, Fagot, Lecerf, & De Ribaupierre, 2009). Hence, inhibitory mechanisms might be important in the avoidance of intrusions. Inhibitory mechanisms might not be as crucial to MOT performance, due to the fact that within the same difficulty level, repeated mentions of the same months were deliberately precluded while constructing the task, and only occur when they become unavoidable, as far apart as possible within each level. Therefore, MOT performance, in addition to a storage component, recruits essentially a manipulation process, with potential intrusions coming mostly from long term storage (i.e., the knowledge of the full 12 months sequence). RST performance has minimal, if any, manipulation demands, but does crucially require inihibition of irrevant information in short term storage. This difference in task requirements perhaps explains the the pattern of correlations with our measures of inhibitory processing.

The MOT and the RST were both moderately and positively correlated with the processing speed measures. The correlations between the RST and the measures of processing speed were higher than those between the MOT and the referred measures. This might be because the RST includes a processing component that explicitly involves coping with a time limit. Participants have only 5 seconds to read each sentence, and the final word must be selected and stored. Indeed, Case and colleagues (1982) found that there is a strong relationship between speed and the number of products that can be stored.

Altogether, the MOT is not highly correlated with measures that assess different (though somewhat related) constructs, except for the DSF.

We also inspected the MOT's reliability. With the purpose of examining the task's temporal stability, 40 participants were assessed on two different occasions (separated by a 12-14 week interval). A 0.925 test-retest coefficient was obtained, surpassing the 0.7 threshold for adequate test-retest results (Nunnally, 1978). In agreement with Anastasi (as cited in Waters & Caplan, 2003), commendable reliability coefficients fall within the 0.80-0.90 range. The result that we have obtained implies that the MOT is capable of yielding consistent scores over time[27]. The RST's test-retest reliability was assessed as well. A 0.640 retest coefficient was attained, a result that casts doubt on the stability of the task. Inappropriate test-retest values for the RST were also reported by other studies (e.g., MacDonald et al., 2001; Waters & Caplan, 1996).

Another type of reliability that was studied concerns the task's internal consistency. Cronbach's alpha was estimated for each administration of the MOT, using the data from the retest sample. The

[27] It is important to keep in mind that we tried to ensure that no confounding variables impacted the participants' retest performance. Testees that underwent significant alterations in terms of their psychological/physical health or personal life that could conceivably affect WM performance were excluded from the retest study. Also, we tried to provide similar testing conditions in both testing sessions.

MOT's alpha coefficient using the entire normative sample was equally calculated. Regarding the internal consistency results for the first administration, a weak alpha value was obtained. The first item was automatically excluded from the analysis due to the high accuracy rates. After the omission of set size three (an item with a low item-total correlation), an acceptable alpha value was attained. The item-total correlations assumed appropriate values, suggesting that the MOT behaved as a homogeneous scale (Ferketich, 1991) in the first part of the test-retest study. The alpha coefficient for the second administration was weak, with set sizes two and three being automatically omitted from the analysis. However, item-total correlations reached once again suitable values, implying that the more demanding set sizes adequately assessed the underlying construct (being and indicator of scale homogeneity). In general, the internal consistency results obtained with the data from the test-retest sample were satisfactory, suggesting that the items of the scale consistently probed WM on different occasions (first test administration and retest), thus ensuring that the temporal stability attested in the test-retest study pertains to a consistent assessment of the target construct, WM. Nevertheless, the automatic exclusion of set size three when assessing the MOT's internal consistency using the retest data suggests the presence of practice effects. These effects might have been influenced by the retest sample's characteristics, as it is quite reduced in size (N=40) and limited in terms of the age range (30-50 years) and years of formal education (M=13.10, SD=4.089), with the initial trials revealing to be too easy for these fairly young and extensively schooled adults. The main interest of the internal consistency study using the test-retest sample's data was to show that the MOT is an internally consistent scale in different points of time and for the test-retest sample in particular. The results using the data of the test-retest participants stress the need to assess the internal consistency of the MOT in a larger sample including a broader age and education range, comprising participants with a reduced number of years of formal education and/or more advanced ages. On that account, we assessed the MOT's internal consistency employing the normative sample's data. Set size two was not automatically excluded from the analysis, suggesting the existence of variability within this item's results. However, when set size two was included in the analysis, a weak alpha value was obtained (α=0.686). After set size two was omitted, the alpha value increased, reaching an acceptable value. The item-total correlations varied once more within the recommended range, supporting the homogeneity of the task.

Even though the initial easier items (set size two) present low item-total correlations and their elimination increases the alpha value, their inclusion in the scale is quite important. Notwithstanding their heavy STM component (which warrants the inadequate item-total correlation), these items play a critical role in the MOT. As previously stated, they might be crucial in the "warming up" of the examinees. Also, we might expect that these initial items could be essential in capturing performance variability within older age groups, as well as within clinical groups.

So as to attain adequate regression-based norms (and to decide which parameters to include in the regression-based algorithm), the influence of demographic characteristics on MOT performance was examined. Age, gender and years of formal education proved to significantly contribute to MOT performance. Only linear relations were observed, as our results did not support the existence of quadraticity in the data. These trends are consistent with the existing literature. Various authors (e.g., Borella et al., 2008; Dobbs & Rule, 1989; Park et al., 2002) have stated that WM performance exhibits continuous and linear age-related declines. It has also been widely observed that WM performance and years of formal education are significantly related, with higher levels of schooling being associated with better WM task scores (e.g., Souza-Talarico, Caramelli, Nitrini, & Chaves, 2007). In terms of gender differences, a variety of studies has reported the existence of dissimilarities for men and women across a vast range of cognitive domains (e.g., Robert & Savoie, 2006).

After examining the influence of demographic variables on MOT performance, regression-based norms for use with speakers of European Portuguese were obtained. Regression-based norms can be seen as "continuous norms" (Zachary & Gorsuch, 1985) in that they do not require the division of the sample in demographic bands, delivering accurate estimates that are derived from the analysis of the entire normative sample (Van der Elst et al., 2012). As said by Maroof (2012), regression-based norms allow an individual's performance to be tailored to his or her unique attributes (i.e., age, gender and education, for instance).

In summary, the MOT revealed to be a valid and reliable instrument. The MOT's correlational trends are quite similar to the RST ones, with both measures showing congruous correlation patterns in terms of convergent and discriminant validity. The fact that the MOT and the RST correlate in an analogous way with other tasks suggest that both measures assess a similar construct. Yet, the obtained results might encourage the use of the MOT instead of the RST. The MOT correlated less strongly with the measures that integrated the discriminant validity study when compared to the RST, which suggests that the MOT measures MT with less sensitivity to peripheral constructs. Also, the MOT presented a stronger correlation with another WM task (DSB) than did the RST, thereby corroborating the idea that the MOT adequately measures WM. Lastly, the RST's retest coefficient assumed an inadequate value, casting doubt on its temporal stability. In contrast, a high retest coefficient for the MOT was obtained. This result suggests that the MOT is a stable instrument over time, producing similar results on different occasions.

VI – Conclusions

WM seems to play a fundamental role in several complex cognitive behaviors, being one of the memory components that are most affected by senescent changes (both normal and pathological). As such, its assessment is important, requiring the existence of adequate WM tasks. Unfortunately, appropriate WM tasks for the older/clinical populations are still quite scarce. Therefore, we intended to study the psychometric properties of the MOT and to obtain normative data for speakers of European Portuguese.

Convergent validity and discriminant validity evidence data for the MOT was obtained, as well as reliability data. Regression-based norms for speakers of European Portuguese were also established. The MOT was shown to be a valid and reliable WM instrument, strongly correlating with measures that assess the same construct and only moderately with instruments that evaluate different (though related) constructs (with the exception of the DSF). The MOT proved to be a temporally stable and internally consistent instrument. Regression analyses revealed that age, gender and years of formal education influence MOT performance. Linear relations were observed between these demographic variables and performance on the stated WM task.

Some suggestions for future studies can be given. It is known that WM tasks tend to significantly correlate with higher order cognitive tasks (such as reasoning, reading comprehension, among others). So as to obtain further convergent validity evidence for this measure, it would be interesting to study the correlations between the MOT and higher order cognitive tasks. An alternative interruption criterion might be proposed, one that would reflect the highest level of difficulty reached by the examinees. As for now, the MOT's final score is heavily influenced by the performance on the initial (easier) items (perhaps explaining the strong correlation with the DSF). The influence of the easier items on the final MOT score might be controlled for by advancing an alternative scoring procedure, with the intent of incrementing the impact, on the final MOT score, of the peak and subsequent decline of the testees' performance. This might more truthfully express the concept of WM, as the peak and ensuing performance degeneration would reflect more closely the ability of the WM system to handle, concomitantly, increasingly exigent manipulation and storage requirements. In an attempt to increase the utility and representativeness of the norms, a broader sample should be attained, especially when it comes to the older age groups. The attainment of a larger sample could also lead to the establishment of traditional norms. One of the strengths of the traditional norms lies within its ease of use, as no additional computations are required and only a normative table has to be consulted. In contrast, regression-based norms are generally not user friendly, as the user often needs to perform some extra calculations.

The MOT seems to possess good psychometric properties and holds certain advantages when compared to other frequently used WM assessment tools. The MOT is quick and easy to administer and has a simple scoring procedure. The instructions are easy to understand and the stimuli relate to the participants' knowledge, possessing ecological validity.

Bibliographical References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30–60.

Almor, A., MacDonald, M. C., Kempler, D., Andersen, E. S., & Tyler, L. K. (2001). Comprehension of long distance number agreement in probable Alzheimer's disease. *Language and Cognitive Processes, 16*(1), 35-63.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, K. L. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language, 67*(3), 202-227.

Bäckman, L., Karlsson, S., Fischer, H., Karlsson, P., Brehmer, Y., Rieckmann, A., . . . Nyberg, L. (2011). Dopamine D1 receptors and age differences in brain activation during working memory. *Neurobiology of Aging, 32*(10), 1849-1856.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology, 8*(2), 234–238.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4*(10), 829-839.

Baddeley, A. (2010). Working Memory. *Current Biology, 4*(23), R136–R140.

Baddeley, A. D., Baddeley, H. A., Bucks, R. S., & Wilcock, G. K. (2001). Attentional control in Alzheimer's disease. *Brain, 124*, 1492-1508.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation, 8*, 47–89.

Baddeley, A., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 38*(4), 603-618.

Ballard, C., Stephens, S., Kenny, R. A., Kalaria, R., Tovee, M., & O'Brien, J. (2003). Profile of neuropsychological deficits in older stroke survivors without dementia. *Dementia and Geriatric Cognitive Disorders, 16*(1), 52-56.

Barrett, T. J. (2011). Computations using analysis of covariance. *Wiley Interdisciplinary Reviews: Computational Statistics, 3*(3), 260–268.

Barreyro, J. P., Burin, D. I., & Duarte, A. D. (2009). Capacidad de la memoria de trabajo verbal: Validez y fiabilidad de una tarea de amplitud de lectura. *Interdisciplinaria, 26*(2), 207-228.

Bastos Leite, A. J., Scheltens, P., & Barkhof, F. (2004). Pathological aging of the brain: An overview. *Topics in Magnetic Resonance Imaging, 15*(6), 369-389.

Beato, R., Levy, R., Pillon, B., Vidal, C., Tezenas du Montcel, S., Deweer, B., . . . Cardoso, F. (2008). Working memory in Parkinson's disease patients. *Arquivos de Neuropsiquiatria*, *66*(2-A), 147-151.

Beckman, B., Holling, H., & Kuhn, J.-T. (2007). Reliability of verbal-numerical working memory tasks. *Personality and Individual Differences, 43*(4), 703-714.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity* (1st ed.). Hoboken, NJ: Wiley.

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica, 128*(1), 33-44.

Borella, E., Ghisletta, P., & de Ribaupierre, A. (2011). Age differences in text processing: The role of working memory, inhibition, and processing speed. *Journal of Gerontology: Psychological Sciences, 66*(3), 311-320.

Buckner, R. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron, 44*(1), 195–208.

Busch, R. M., Chelune, G. J., & Suchy, Y. (2006). Using norms in neuropsychological assessment of the elderly. In D. K. Attix, & K. A. Welsch-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 133-157). New York, NY: Guilford Press.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*(1), 85-100.

Calderon, J., Perry, R. J., Erzinclioglu, S. W., Berrios, G. E., Dening, T. R., & Hodges, J. R. (2001). Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 70*, 157-164.

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39*(2), 231–244.

Carlesimo, G. A., & Oscar-Berman, M. (1992). Memory deficits in Alzheimer's patients: A comprehensive review. *Neuropsychology Review, 3*(2), 119-169.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psycholog, 33*(3), 386-404.

Castro, S. L., Cunha, L. S., & Martins, L. (2000). *Teste Stroop Neuropsicológico em português*. Retrieved from http://www.fpce.up.pt/labfala

Cipolotti, L., & Warrington, E. K. (1995). Neuropsychological assessment. *Journal of Neurology, Neurosurgery, and Psychiatry, 58*(6), 655-664.

Coalson, D., Raiford, S. E., Saklofske, D. H., & Weiss, L. G. (2010). WAIS-IV: Advanced in the assessment of intelligence. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical* *use and interpretation: Scientist-practitioner perspectives* (pp. 3-20). San Diego, CA: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connor, L. T. (2001). Memory in old age: Patterns of decline and preservation. *Seminars in Speech and Language, 22*(2), 117-125.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163-191.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2005). Working-memory capacity limits in a theoretical context. In C. Izawa, & N. Ohta (Eds.), *Human learning and memory: Advances in theory and applications. The 4th Tsukuba international conference on memory* (pp. 155-175). Mahwah, NJ: Erlbaum.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. Sossin, J.-C. Lacaille, V. Catellucci, & S. Belleville (Eds.), *Progress in Brain Research: Essence of Memory* (Vol. 169, pp. 323-338). Amsterdam, the Netherlands: Elsevier.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour, 19*(4), 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561-584.

DeDe, G., Ricca, M., Knilans, J., & Trubl, B. (2014). Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology, 28*(6), 692-712.

Dehn, M. (2008). *Working memory and academic learning: Assessment and intervention*. Hoboken, NJ: Wiley.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London B. Biological Sciences, 362*, 761–772.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain and Cognition, 41*(1), 66-86.

D'Esposito, M., Postle, B. R., Jonides, J., & Smith, E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 96*(13), 7514–7519.

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4*(4), 500-503.

Drozdick, L. W., Holdnack, J. A., & Hilsabeck, R. C. (2011). *Essentials of WMS-IV assessment.* New York, NY: Wiley.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Ferketich, S. (1991). Focus on psychometrics: Aspects of item analysis. *Research in Nursing & Health, 14*(2), 165-168.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Foos, P. W. (1995). Working memory resource allocation by young, middle-aged, and old adults. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 21*(3), 239-250.

Freitas, S., Simões, M. R., Martins, C., Vilar, M., & Santana, I. (2010). Estudos de adaptação do Montreal Cognitive Assessment (MoCA) para a população portuguesa. *Avaliação Psicológica, 9*(3), 345-357.

Fuster, J. M., & Garrett, A. E. (1971). Neuron activity related to short-term memory. *Science, 173*(3997), 652-654.

Gabrieli, J. D., Singh, J., Stebbins, G. T., & Goetz, C. G. (1996). Reduced working memory span in Parkinson's disease: Evidence for the role of a frontostriatal system in working and strategic memory. *Neuropsychology, 10*(3), 322-332.

Gold, J., Carpenter, C., Randolph, C., Goldberg, T., & Weinberger, D. (1997). Auditory working memory and the Wisconsin Card Sorting Test. *Archives of General Psychiatry, 54*(2), 159-165.

Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America, 93*(24), 13473-13480.

Goral, M., Clark-Cotton, M., Spiro, A., Obler, L. K., Verkuilen, J., & Albert, M. L. (2011). The contribution of set switching and working memory to sentence processing in older adults. *Experimental Aging Research, 37*(5), 516–538.

Greenwood, P. M. (2000). The frontal aging hypothesis evaluated. *Journal of the International Neuropsychological Society, 6*(6), 705–726.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.).

Hoboken, NJ: Wiley.

Harkin, B., Rutherford, H., & Kessler, K. (2011). Impaired executive functioning in subclinical compulsive checking with ecologically valid stimuli in a working memory task. *Frontiers in Psychology, 2*(78), 1-10.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rympa, B. (1991). Age and inhibition. *Journal of Experimental Psychology Learning, Memory, and Cognition, 17*(1), 163-169.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation, 22*, 193–225.

Haut, M. W., Kuwabara, H., Leach, S., & Arias, R. G. (2000). Neural activation during performance of number-letter sequencing. *Applied Neuropsychology, 7*(4), 237-242.

Hill, B. D. (2008). *The construct validity of the clinical assessment of working memory ability* (Unpublished master's thesis). Louisiana State University, Baton Rouge, Louisiana, USA.

Hill, B. D., Elliott, E., Shelton, J., Pella, R., O'Jile, J., & Gouvier, W. (2010). Can we improve the clinical assessment of working memory? An evaluation of the WAIS-III using a working memory criterion construct. *Journal of Clinical and Experimental Neuropsychology, 32*(3), 315–323.

Hockey, A., & Geffen, G. (2004). The concurrent validity and test–retest reliability of a visuospatial working memory task. *Intelligence, 32*(6), 591–605.

Hoyer, W. J., & Verhaeghen, P. (2006). Memory aging. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 209-232). San Diego, CA: Elsevier.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *n*-back task as a working memory measure. *Memory, 18*(4), 394-412.

Jaiswal, N., Ray, W., & Slobounov, S. (2010). Encoding of visuo-spatial information in working memory requires more cerebral efforts than retrieval: Evidence from EEG and virtual reality study. *Brain Research, 1347*, 80–89.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139*(1), 39–50.

Jo, J., Kim, Y., Ko, M. , Ohn, S. H., Joen, B., & Lee, K. H. (2009). Enhancing the working memory of stroke patients using tDCS. *American Journal of Physical Medicine & Rehabilitation, 88*(5), 404-409.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169-183.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition, 33*(3), 615-622.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology, 133*(2), 189-217.

Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. (1998). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language, 64*(3), 297-316.

Kensinger, E. A., Shearer, D. K., Locascio, J. J., Growdon, J. H., & Corkin, S. (2003). Working memory in mild Alzheimer's disease and early Parkinson's disease. *Neuropsychology, 17*(2), 230–239.

Kirshner, H., & Mark, V. (2009). Ischemic and Intracerebral Hemorragic Stroke. In J. R. Festa, & R. M. Lazar (Eds.), *Neurovascular neuropsychology* (pp. 19-48). New York, NY: Springer.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352-358.

Klaassen, E. B., Evers, E. A., De Groot, R. H., Backes, W. H., Veltman, D. J., & Jolles, J. (2014). Working memory in middle-aged males: Age-related brain activation changes and cognitive fatigue effects. *Biological Psychology, 96*, 134–143.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments & Computers, 31*(3), 429-432.

Kljajevic, V., Fratini, V., Etxaniz, A., Urdaneta, E., & Yanguas, J. (2013). Comprehension cueing strategies in elderly: A window into cognitive decline? *Cognitive Science Society Proceedings, 32*, 2746-2751.

Kolb, B., & Whishaw, I. Q. (2009). *Fundamentals of human neuropsychology* (6th ed.). New York, NY: Worth Publishers.

Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. (2001). Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology, 19*(3), 339–348.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment.* New York, NY: Wiley.

Linden, D. E. (2007). The working memory networks of the human brain. *Neuroscientist, 13*(3), 257-267.

Luo, L., & Craik, F. I. (2008). Aging and memory: A cognitive approach.

Canadian Journal of Psychiatry, 53(6), 346–353.

MacDonald, M. C., Almor, A., Henderson, V. W., Kempler, D., & Andersen, E. S. (2001). Assessing working memory and language comprehension in Alzheimer's disease. *Brain and Language, 78*(1), 17-42.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

Manning, C. A., & Ducharme, J. K. (2010). Dementia syndromes in the older adult. In P. A. Lichtenberg (Ed.), *Handbook of assessment in clinical gerontology* (2nd ed., pp. 155-178). San Diego, CA: Academic Press.

Maroof, D. A. (2012). *Statistical methods in neuropsychology: Common procedures made comprehensible.* New York, NY: Springer.

Marquardt, D. W. (1980). You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association, 75*(369), 87-91.

Maruff, P., & Darby, D. (2006). Age-related memory impairment. In P. J. Snyder, P. D. Nussbaum, & D. Robins (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (2nd ed., pp. 155–182). Washington, DC: American Psychological Association.

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., . . . Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters, 392*(1-2), 32–37.

Mattson, M. P., & Magnus, T. (2006). Ageing and neuronal vulnerability. *Nature Reviews Neuroscience, 7*, 278-294.

McAllister, T. W., Flashman, L. A., McDonald, B. C., & Saykin, A. J. (2006). Mechanisms of working memory dysfunction after mild and moderate TBI: Evidence from functional MRI and neurogenetics. *Journal of Neurotrauma, 23*(10), 1450-1467.

McAllister, T. W., Flashman, L. A., Sparling, M. B., & Saykin, A. J. (2004). Working memory deficits after traumatic brain injury: Catecholaminergic mechanisms and prospects for treatment - a review. *Brain Injury, 18*(4), 331-350.

McAuley, T., & White, D. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology, 108*(3), 453–468.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Morey, C. C., Cowan, N., Morey, R. D., & Rouder, J. N. (2011). Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a trade-off. *Perception & Psychophysics, 73*(2), 458-472.

Morgado, J., Rocha, C. S., Maruta, C., Guerreiro, M., & Martins, I. P. (2009). Novos valores normativos do Mini-Mental State

Examination. *Sinapse, 9*(2), 10-16.

Morris, R. G., Downes, J. J., Sahakian, B. J., Evenden, J. L., Heald, A., & Robbins, T. W. (1988). Planning and spatial working memory in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 51*(6), 757-766.

Morris, T. (2010). Traumatic brain injury. In C. L. Armstrong, & L. Morrow (Eds.), *Handbook of medical neuropsychology: Applications of cognitive neuroscience* (pp. 17-32). New York, NY: Springer.

Moseley, M. (2002). Diffusion tensor imaging and aging: A review. *NMR in Biomedicine, 15*(7-8), 553-560.

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

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Oberauer, K., Süss, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity: Facets of a cognitive ability construct. *Personality and Individual differences, 29*(6), 1017-1045.

Oberauer, K., Süss, H. M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167-193.

Olkkonen, S. (2013). Speed in cognitive tasks as an indicator of second/foreign language reading and writing skills. *Estonian Papers in Applied Linguistics, 9*, 195–208.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*(2), 299-320.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Pestana, M. H., & Gageiro, J. N. (2008). *Análise de dados para ciências sociais: A complementaridade do SPSS* (3rd ed.). Lisbon, Portugal: Edições Sílabo.

Potter, G. G., & Attix, D. K. (2006). An integrated model for geriatric neuropsychological assessment. In D. K. Attix, & K. A. Welsh-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 5-26). New York, NY: Guilford Press.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience, 3*(1), 85–90.

Redick, T. S., & Lindsey, D. R. (2013). Complex span and *n*-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20*(6), 1102-1113.

Reuter-Lorenz, P. A. (2002). New visions of the aging mind and brain. *Trends in Cognitive Sciences, 6*(9), 394–400.

Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C., & Koeppe, R. A. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience, 12*(1), 174-187.

Reuter-Lorenz, P. A., Stanczak, L., & Miller, A. C. (1999). Neural recruitment and cognitive aging: Two hemispheres are better than one, especially as you age. *Psychological Science, 10*(6), 494-500.

Reynolds, C. R. (1997). Forward and Backward memory Span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12*(1), 29-40.

Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 573–585.

Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working memory resources? *European Journal of Cognitive Psychology, 18*(3), 378-397.

Robert, C., Borella, E., Fagot, D., Lecerf, T., & de Ribaupierre, A. (2009). Working memory and inhibitory control across the life span: Intrusion errors in the Reading Span Test. *Memory & Cognition, 37*(3), 336-345.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*(6), 573-597.

Rouleau, N., & Belleville, S. (1996). Irrelevant speech effect in aging: An assessment of inhibitory processes in working memory. *Journal of Gerontology: Psychological Sciences, 51B*(6), 356-363.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763-776.

Sander, M. C., Lindenberger, U., & Werkle-Bergner, M. (2012). Lifespan age differences in working memory: A two-component framework. *Neuroscience and Biobehavioral Reviews, 36*(9), 2007–2033.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1089-1096.

Schneider-Garces, N. J., Gordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., . . . Fabiani, M. (2009). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience, 22*(4), 655–669.

Schulze, E. T., Geary, E. K., Susmaras, T. M., Paliga, J. T., Maki, P. M., & Little, D. M. (2011). Anatomical correlates of age-related working memory declines. *Journal of Aging Research, 2011*, 1-9.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology, 125*(1), 4-27.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*(1), 11-20.

Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Smith, G., & Rush, B. K. (2006). Normal aging and mild cognitive impairment. In D. K. Attix, & K. A. Welsh-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 27-55). New York, NY: Guilford Press.

Sokal, R. R., & Rohlf, J. F. (1995). *Biometry: The principles and practices of statistics in biological research* (3rd ed.). New York, NY: W.H. Freeman.

Souza-Talarico, J., Caramelli, P., Nitrini, R., & Chaves, E. (2007). The influence of scholling on working memory performance in elderly individuals without cognitive decline. *Dementia & Neuropsychologia, 1*(3), 276-281.

Spaan, P. E., Raaijmakers, J. G., & Jonker, C. (2005). Early assessment of dementia: The contribution of different memory components. *Neuropsychology, 19*(5), 629–640.

Spores, J. M. (2013). *Clinician's guide to psychological assessment and testing: With forms and templates for effective testing.* New York, NY: Springer.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). Memory. In E. Strauss, E. M. S. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 678-890). New York, NY: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–661.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1985). *The Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.

Turner, M. R., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154.

Van Breukelen, G. J., & Vlaeyen, J. W. (2005). Norming clinical questionnaires with multiple regression: The pain cognition list. *Psychological Assessment, 17*(3), 336–344.

Van der Elst, W., Dekker, S., Hurks, P., & Jolles, J. (2012). The Letter Digit Substitution test: Demographic influences and regression-based normative data for school-aged children. *Archives of Clinical Neuropsychology, 27*(4), 433-439.

Wang, M., Gamo, N. J., Yang, Y., Jin, L. E., Wang, X.-J., Laubach, M., . . . Arnsten, A. F. (2011). Neuronal basis of age-related working memory decline. *Nature, 476*(7359), 210-213.

Ward, C. L., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*(2), 328–340.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A*(1), 51-79.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behaviour Research Methods,* *Instruments, & Computers*, *35*(4), 550-564.

Wechsler, D. (2008). *Escala de Memória de Wechsler* (3rd ed.). Lisbon, Portugal: CEGOC-TEA.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale* (4th ed.). San Antonio, TX: Pearson.

Whatmough, C. (2010). Dementia. In C. L. Armstrong, & L. Morrow (Eds.), *Handbook of medical neuropsychology: Applications of cognitive neuroscience* (pp. 277-295). New York, NY: Springer.

Wilson, F. A., Óscalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, *260*(5116), 1955-1958.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? J*ournal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255-260.

Wright, H. H., & Fergadiotis, G. (2012). Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology*, *26*(3-4), 258-278.

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*(1), 86-94.

Appendix I – Correlation matrix for measures of memory, inhibition and processing speed.

	1	2	3	4	5	6	7	8	9	10
Memory tasks										
1. MOT	-	0.589**	0.617**	0.513**	-0.032	-0.210	0.254*	-0.021	0.353**	0.342**
2. Reading Span Task	-	-	0.395**	0.402**	0.081	-0.259	0.305**	0.099	0.429**	0.426**
3. Digit Span Forward	-	-	-	0.512**	-0.029	-0.314*	0.339*	0.040	0.395**	0.387**
4. Digit Span Backward	-	-	-	-	-0.030	-0.340*	0.391**	0.209	0.286*	0.281
Inhibitory efficiency measures										
5. RCWpc-RBWpc	-	-	-	-	-	-0.064	-0.042	-0.203	0.254*	0.246*
6. RCWs-RBWs	-	-	-	-	-	-	0.011	0.100	-0.166	-0.165
7. NCWpc-NCSpc	-	-	-	-	-	-	-	0.302**	0.083	0.051
8. NCWs-NCSs	-	-	-	-	-	-	-	-	-0.423**	-0.415**
Measures of processing speed										
9. NCSs	-	-	-	-	-	-	-	-	-	0.997**
10. NCSsr	-	-	-	-	-	-	-	-	-	-

*p<0.05, **p<0.01

Appendix II – Hierarchical multiple regression analysis assessing the contribution of demographic variables (age, gender, years of formal education) to MOT performance.

	B	*SE*	*Β*
Step 1			
Age (centered)	-0.245	0.029	-0.498***
Step 2			
Age (centered)	-0.266	0.029	-0.521***
Age^2	-0.006	0.002	-0.175**
Step 3			
Age (centered)	-0.289	0.030	-0.567***
Age^2	-0.004	0.002	-0.125
Gender	3.365	1.195	0.173**
Step 4			
Age (centered)	-0.173	0.034	-0.339***
Age^2	-0.000022	0.002	-0.001
Gender	2.551	1.120	0.131*
Education (centered)	0.988	0.165	0.395***
Step 5			
Age (centered)	-0.201	0.037	-0.394***
Age^2	0.000004	0.002	0.000
Gender	2.480	1.115	0.127*
Education (centered)	0.992	0.164	0.397***
$Education^2$	0.060	0.032	0.112

Notes. R^2=0.248 for step 1, ΔR^2= 0.030 (p<0.01) for step 2, ΔR^2=0.025 (p<0.01) for step 3, ΔR^2=0.096 (p<0.001) for step 4, ΔR^2=0.009 (p=0.062) for step 5. R^2=R squared, ΔR^2=R squared change, *B*=unstandardized regression coefficient, *SE*=standard error, *β*=standardized regression coefficient. *p<0.05, **p<0.01, ***p<0.001

Appendix III - Two-way between-groups analysis of covariance (examination of the impact of age group on MOT performance).

Statistical Analysis

So as to further explore the impact of age on MOT scores and to examine the possible differences in MOT performance between the various age groups, a two-way between-groups analysis of covariance was conducted (controlling for the influence of other demographic variables: gender, which entered the analysis as a fixed factor, and years of formal education, which entered the analysis as a covariate). The dependent variable was the proportion-correct MOT score for each testee (taking into consideration the interruption criterion), and the independent variable was the participants' age group. The continuous age variable was firstly categorized by quartiles. With the purpose of obtaining meaningful age groups (i.e., within which no significant effect of age upon MOT performance would subsist), linear regression analyses were performed within each quartile so as to determine whether the age variable revealed to be a significant predictor of MOT performance. Whenever age was found to have a significant effect on MOT performance, the quartile was split into two (based on the median), and linear regression analyses were repeated until age ceased to be a significant predictor. This procedure was carried out for each age quartile, and the following age groups were attained: 18-20 years, 21-30 years, 31-46 years, 47-64 years, 65-71 years and 72-80 years. The Bonferroni correction was used, and the assumptions that underlie the ANCOVA were tested (normality[28], linearity, homogeneity of variances and homogeneity of regression slopes).

For this study we used the data from the normative sample. After the exclusion of one participant due to a low MMSE score, the sample comprised 247 participants, with ages ranging between 18 and 80 years (M=46.664, SD=19.073).

Results

The continuous age variable was divided into six age categories. Within each one of these groups age did not have a significant influence on MOT performance. The analysis revealed a significant effect of age group regarding MOT performance [$F(5, 239)=3.950$, $p<0.01$, partial $\eta^2=0.076$]. Age explains 7.6% of the variance in the MOT scores.

Table 1. Between-subjects summary.

Source	Sum of Squares	df	Mean square	*F*	Partial η^2
Age	1793.303	1	358.661	3.950**	0.076
Gender	160.578	1	160.578	1.768	0.007

[28] The dependent variable did not meet the assumption of normality. Nevertheless, it is argued that ANCOVA is robust to violations of normality (e.g., Barrett, 2011). Furthermore, the MOT scores were normally distributed within each age group.

Error	21701.583	239	90.801

Notes. df=degrees of freedom, Partial η^2=partial eta-squared. **p<0.01

Examining these results in more depth (by means of Bonferroni corrected univariate comparisons) (table 2), it is possible to conclude that the scores obtained by the 72-80 age group (*M*=0.487, *SD*=0.131) are significantly different from those attained by the 18-20 age group (*M*=0.722, *SD*=0.104), 21-30 age group (*M*=0.756, *SD*=0.114) and 31-46 age group (*M*=0.709, *SD*=0.151). No significant differences were found between the MOT scores of the 72-80 age group and the 47-64 age group (*M*=0.642, *SD*=0.177), and neither among the 72-80 age group and the 65-71 age group (*M*=0.563, *SD*=0.146). No statistically significant differences were observed amid other age groups.

Table 2. Descriptive statistics (using untransformed proportion-correct MOT scores, considering the interruption criterion) for age categories, and pairwise comparisons between groups (mean differences).

Age groups	*M* (*SD*)	*n*	1	2	3	4	5	6
1. 18-20	0.722 (0.104)	39	-	-0.034	0.013	0.080	0.159	0.235**
2. 21-30	0.756 (0.114)	24	-	-	0.047	0.114	0.193	0.269**
3. 31-46	0.709 (0.151)	61	-	-	-	0.067	0.146	0.222*
4. 47-64	0.642 (0.177)	65	-	-	-	-	0.079	0.155
5. 65-71	0.563 (0.146)	31	-	-	-	-	-	0.076
6. 72-80	0.487 (0.131)	27	-	-	-	-	-	-

Notes. M=mean, *SD*=standard deviation, *n*=sample size. *p<0.05, **p<0.01

Discussion

In order to better understand how age affects MOT performance throughout the several age groups, an additional analysis was carried out. A two-way between-groups analysis of covariance was employed, and significant differences in MOT scores were observed between the oldest group (72-80 years of age) and the younger groups (18-20 years of age, 21-30 years of age, 31-46 years of age), with the older group exhibiting lower scores than the younger groups. The results suggest that WM declines with age in a characteristic pattern and, in our sample, these declines seemed to become most evident from the seventy-second year of life onwards. As stated by Connor (2001, p. 119), “declines in WM in individuals 70 years of age and older are profound (…)”. Even though the MOT scores declined throughout the different age groups, no significant differences between other groups were found. In accordance with other studies (e.g., Foos, 1995, Klaassen et al., 2014), the performance of our “middle-age” and “young-elderly” groups (47-64 years, 65-71 years) did not significantly differ from the performance of the younger adults, but neither did it differ from that of the “old-elderly” group. Apparently, the decline in these intermediate age ranges is substantial enough to blur the statistical significance of the comparison with the “old-elderly”, but also subtle enough to blur its distinctiveness with respect to the results of the younger groups, namely the 21-30 group, in which WM performance seems to be in its prime. The MOT seemed to be sensitive to age differences, distinguishing the performance of older and younger participants.

As the MOT appeared to be sensitive to WM declines due to normative senescent processes, it might be possibly useful in identifying WM differences due to pathological processes. It would be applicable to carry out studies with patients suffering from relevant aging-associated disorders. As mentioned in the theoretical framework, Kljajevic and colleagues (2013) found that the MOT was able to differentiate WM performance of healthy eldery participants and participants with mild cognitive impairment. Some studies demonstrated that this task is adequate for the assessment of demented populations’ WM (Almor et al., 1999, 2001; MacDonald et al., 2001).