



UC/FPCE_2012

Universidade de Coimbra
Faculdade de Psicologia e de Ciências da Educação

Testes de Validade de Sintomas: Validação de um protocolo em amostras de adultos idosos

Joana Isabel de Sousa Santos Pinho
(*e-mail*: ju_pinho@hotmail.com)

Dissertação de Mestrado Integrado em Psicologia, **área de especialização em Psicologia Clínica e da Saúde**, subárea de especialização em **Psicogerontologia Clínica**, sob a orientação de **Professor Doutor Mário R. Simões** (Prof. Catedrático da FPCE-UC) e co-orientação da **Dr.ª Maria Fernanda dos Prazeres Duarte** (Psicóloga Clínica, Mestre em Psicologia, CHUC -Unidade Sobral Cid).

“Any neuropsychological evaluation that does not include careful consideration of the patient’s motivation to give their best effort should be considered incomplete”
(Iverson, 2003)

“The assessment of symptom validity is a crucial component to neuropsychological assessment, and clinicians should be prepared to justify not assessing symptom validity as part of the neuropsychological examination”
(Bush et al., 2005)

“The use of one test alone is normally strongly discouraged when assessing the presence or absence of suboptimal effort”
(Rudman, Oyeboode, Jones, & Bentham, 2011, p. 56)

Testes de Validade de Sintomas: Validação de um protocolo em amostras de adultos idosos

A avaliação da *simulação, exagero de sintomas* ou *esforço insuficiente* ou *reduzido* tem recebido atenção crescente. A preocupação com a possibilidade de exagero de sintomas aquando das avaliações neuropsicológicas é actualmente uma questão de reconhecimento relevante. Com a proliferação de instrumentos e investigação nesta área surge a questão de saber que instrumentos utilizar bem como a possibilidade de obtenção de dados contraditórios nas diferentes medidas utilizadas num protocolo de avaliação.

O objectivo primordial deste trabalho consiste na validação de um protocolo de testes de esforço reduzido ou insuficiente em contexto de avaliação neuropsicológica. Procedeu-se à aplicação de diferentes medidas de validação de desempenho num grupo da comunidade (n=31), com Declínio Cognitivo Ligeiro (n=17) e num grupo instruído a "simular défices cognitivos" (n=29).

O protocolo utilizado na presente investigação inclui os seguintes instrumentos: questionário de dados demográficos, *Addenbrooke's Cognitive Examination*, Escala de Depressão Geriátrica, Inventário de Avaliação Funcional de Adultos Idosos, Inventário de Ansiedade Geriátrica, subtteste Memória Lógica I e II da WMS-III, *Test of Memory Malingering*, subtteste Memória de Dígitos da WAIS-III, *b-Test*, subtteste Vocabulário da WAIS-III, *Dot Counting Test*, *Structured Inventory of Malingered Symptomatology*, e *Rey 15-Item Memory Test*.

Os resultados do presente estudo permitem concluir que, contrariamente ao pressuposto básico dos testes de esforço reduzido, algumas medidas utilizadas são sensíveis ao declínio, bem como idade e escolaridade, o que evidencia a necessidade de prudência acrescida na interpretação dos resultados em amostras clínicas e a consequente necessidade de ajuste dos pontos de corte para estas populações. No grupo instruído a simular, os sujeitos não são identificados todos pelas mesmas medidas, o que evidencia a utilidade do recurso a diferentes instrumentos. Os valores de especificidade elevados na maioria das medidas aplicadas reforçam a validade e utilidade do presente protocolo. O Ensaio 2 e Ensaio de Reconhecimento do TOMM, Índice de Respostas Infrequentes do subtteste

Memória Lógica, pontuação total da Memória de Dígitos, Ensaio de Evocação Imediata (<6) do 15-IMT e o índice Vocabulário menos Memória de Dígitos são os que aparentam ser mais indicados nestas amostras, com valores de especificidade superiores a 85%. Apesar dos valores de sensibilidade baixos, as taxas de falsos positivos reduzidas são preferíveis nestes casos, sob pena de considerar erradamente indivíduos como estando a simular ou exagerar sintomas. O DCT foi o instrumento que aparentemente apresenta menor validade incremental ao presente protocolo uma vez que apresenta 100% de falsos negativos.

Palavras-chave: simulação, esforço insuficiente, testes de validade de sintomas, avaliação neuropsicológica, adultos idosos

Symptom Validity Tests: Validation of a protocol in older adults samples

The assessment of malingering, insufficient or poor effort has received a great deal of attention. The concern with the possibility of symptom exaggeration in neuropsychological evaluations is currently a question of relevant recognition. With the proliferation of instruments and concomitant investigation in this area emerge the question of know which instruments use, as well the possibility of obtaining conflicting results in the different measures used in an assessment protocol.

The main goal of this work is the validation of a protocol of Symptom Validity Tests in neuropsychological assessment context. We apply different measures of performance validity in a community group (n=31), with mild cognitive impairment (n=17) and in a group instructed to feign (n=29).

The protocol used in the presente investigation includes the following instruments: demographics questionnaire, Addenbrooke's Cognitive Examination, Escala de Depressão Geriátrica, Inventário de Avaliação Funcional de Adultos Idosos, Inventário de Ansiedade Geriátrica, subtest Memória Lógica I and II of WMS-III, *Test of Memory Malingering*, subtest Memória de Dígitos of WAIS-III, *b-Test*, subtest Vocabulário of WAIS-III, *Dot Counting Test*, *Structured Inventory of Malingered Symptomatology*, and *Rey 15-Item Memory Test*.

The results of the present study show that, contrary to the initial assumption, some measures are sensible to decline, and also age and education, what highlights the need of increased caution in the interpretation of results in clinic samples and consequent need of cut-off adjustments in these populations. In the group instructed to feign, the subjects are not all identified in the same measures, what shows the utility of using different instruments. The high values of specificity in the majority measures applied reinforce the validity and utility of present protocol. The Trial 2 and the Retention Trial of TOMM, Rarely Missed Index of Logical Memory, score of Digit Span, free recall score of 15-IMT (<6) and the index Vocabulary minus Digit Span are the measures that apparently are more indicated in this samples, with specificity values over 85%. Despite the lower values of sensibility, the low rates of false positives are better, otherwise consider wrong subjects like being feigning or exaggerate symptoms. The DCT was

the instrument that apparently adds little incremental validity to the present protocol with rates of 100% false negatives.

Key Words: malingering, insufficient effort, symptom validity tests, neuropsychological evaluation, older adults

Agradecimentos

Ao Professor Doutor Mário R. Simões, pelo exemplo de competência e sabedoria transmitidos, pela liberdade dada para seguir as minhas ideias, ainda que com supervisão, correcções, directrizes e sugestões permanentes. Obrigada pela disponibilidade e ensinamentos contantes; apesar do horário sempre demasiado preenchido, o apoio e supervisão não faltaram.

À Dr.^a Fernanda Duarte, por me chamar à razão quando precisei, por me permitir a recolha de dados e pela compreensão nos momentos mais difíceis.

Aos meus pais, por estarem sempre presentes, cada um à sua maneira, e por esforçarem-se por entender e aceitar as minhas escolhas. Por todo o amor e carinho despendido, pela força e coragem que me incutiram quando a vontade de facto era deixar tudo para trás, por não me deixarem desistir face às dificuldades e por me acompanharem durante todo este percurso.

A ti, Jorge, companheiro das horas que, boas ou más, passaram rápido demais neste ano em que o tempo é sempre escasso. Obrigado pelo apoio e interesse incondicional, por não me deixares desistir deste meu percurso e por me fazeres sempre acreditar que é possível “com calma, uma coisa de cada vez”...

À minha madrinha, amiga e cunhada... pela ajuda incondicional e indefinível, pela preocupação e incentivos e, sobretudo, pelo sacrifício realizado para me ajudar sempre que necessitei.

A todos os meus familiares, e em especial aos meus avós paternos, por sempre terem acreditado em mim. Ao meu avô que, apesar de não me ter visto a terminar esta fase, sempre acreditou que conseguisse e sempre se mostrou orgulhoso de todo o meu percurso. À minha avó não consigo transcrever por palavras a gratidão que tenho. Por todas as viagens realizadas a Coimbra, pela companhia, pelos momentos partilhados, pelos conselhos, pela amizade, pela confiança, pelos incentivos para continuar a lutar quando tudo parece não ter solução... muito obrigada!

À Lurdes, uma amiga para toda a vida. Obrigada pela amizade, disponibilidade, paciência interminável para ouvir todos os meus receios e pelos momentos partilhados (e também pelos jantares de “descontração” que nos faziam tão bem após as nossas reuniões de estágio).

Um especial agradecimento a todos os sujeitos que aceitaram participar no presente estudo, pela disponibilidade que demonstraram.

A todos aqueles que não me ajudaram e não apoiaram e a todas as instituições que não deram o seu contributo, por me fazerem perceber que não existem apenas facilidades e me alertarem para os obstáculos e dificuldades que certamente irei enfrentar ao longo do meu percurso. Todos estes “nãos” com certeza me tornaram menos ingénua face à utopia de que tudo correrá como planeamos inicialmente. De facto, foram bastantes os ajustes que foram sendo necessários para se tornar viável a concretização e finalização deste projecto.

Índice

Introdução.....	1
I – Enquadramento conceptual	1
1. Definição de simulação	2
2. Abordagens na detecção de simulação.....	5
3. Testes de validade de sintomas usados na detecção de simulação ou esforço insuficiente de défices cognitivos.....	7
3.1 TOMM.....	7
3.2 <i>Dot Counting Test</i>	10
3.3 <i>Rey 15-Item Memory Test (15-IMT)</i>	11
3.4 <i>b-test</i>	12
4. Testes de Validade de Sintomas usados na detecção de simulação ou esforço insuficiente de défices psicopatológicos	14
4.1 <i>Structured Inventory of Malingered Symptomatology</i>	14
5. Outras medidas usadas na detecção de simulação e/ou esforço insuficiente.....	15
5.1 Subteste Memória de Dígitos e Memória Lógica.....	15
6. Condições Clínicas: Demência, Depressão e ansiedade	16
II - Objectivos.....	19
III - Metodologia	20
1. Amostra.....	20
2. Instrumentos.....	22
2.1 Questionário de dados demográficos.....	22
2.2 <i>Addenbrooke's Cognitive Examination</i>	22
2.3 Inventário de Avaliação Funcional de Adultos Idosos.....	22
2.4 <i>Test of Memory Malingered</i>	23
2.5 <i>Dot Counting Test</i>	23
2.6 <i>b-Test</i>	23
2.7 <i>Rey 15-Item Memory Test</i>	24
2.8 <i>Structured Inventory of Malingered Symptomatology</i>	24
2.9 Subteste Vocabulário.....	24
2.10 Subteste Memória de Dígitos.....	25
2.11 Subteste Memória Lógica I e II.....	25

2.12 Escala de Depressão Geriátrica – 30 itens	25
2.13 <i>Geriatric Anxiety Inventory</i>	25
3. Procedimentos.....	25
IV - Resultados.....	26
1. Comparação do desempenho dos grupos nas diferentes medidas.....	26
2. Efeito das variáveis idade, escolaridade e género no desempenho nos testes de simulação e/ou esforço insuficiente (para cada um dos grupos)	28
3. Análise da influência das alterações de humor nos desempenhos nos diferentes resultados dos TVS.....	31
4. Análise dos valores de especificidade (grupo da comunidade e DCL) e sensibilidade (instrução) dos diferentes índices	31
5. Análise crítica dos pontos de corte propostos para o Ensaio 1 do TOMM.....	33
6. Contributo de cada medida na identificação de simulação ou esforço insuficiente	35
V - Discussão.....	37
VI - Conclusões.....	43
Bibliografia.....	44
Anexos.....	51
Anexo 1: Critérios de diagnóstico de Simulação de Perturbação Neurocognitiva (SPN) (Slick et al., 1999)	52
Anexo 2: Itens do Índice de Respostas Infrequentes e respectiva ponderação	53
Anexo 3: Instrução dada à amostra da comunidade instruída a simular ..	54
Anexo 4: Análise comparativa do desempenho no grupo com DCL amnésico e não amnésico.....	55
Anexo 5: Análise dos valores de correlação entre depressão e ansiedade e os resultados do SIMS (amostra da comunidade e DCL)	56
Anexo 6: Médias e desvios padrão dos desempenhos nas diferentes medidas nos grupos Comunidade e DCL.....	57

Introdução

A validade dos sintomas cognitivos reportados pelos sujeitos é habitualmente corroborada ou refutada pelos resultados nos testes. Contudo, os resultados podem estar comprometidos por outros motivos que não os défices cognitivos reais (ou presença de sintomas psicopatológicos), como motivação reduzida, esforço insuficiente, simulação ou exagero de sintomas (Simões et al., 2010; Jasinski, Berry, Shandera, & Clark, 2011). Este tipo de respostas é frequente em situações de litígio em que os ganhos secundários podem justificar a tendência dos sujeitos para simular ou exagerar sintomas nas avaliações efectuadas (ex. determinação de incapacidade e recompensas). Neste contexto, o recurso a Testes de Validade de Sintomas (TVS) torna-se indispensável e útil para avaliar a validade dos resultados dos testes (cf. Simões et al., 2010; Bliger, 2012). Apesar de reconhecida a importância dos TVS em situações que não apenas de litígio, o seu uso ainda não é sistemático em contexto de avaliação neuropsicológica (cf. O'Bryant, Engel, Kleiner, Vasterling, & Black, 2007; Simões et al., 2010). Todavia, tal como referido por Iverson (2003, cit. in Bauer, O'Bryant, Lynch, Mccaffrey & Fisher, 2007) as avaliações neuropsicológicas que não incluam a determinação do grau de motivação dos sujeitos devem ser consideradas incompletas. Bush e colaboradores (2005) defendem mesmo que os profissionais devem justificar a não avaliação da validade dos desempenhos aquando das avaliações neuropsicológicas.

Na população idosa também se podem identificar situações onde os ganhos secundários estão presentes e expliquem a eventual simulação de défices nos diferentes TVS. A título de exemplo refere-se os casos de reformas antecipadas, exposição a tóxicos, acesso a fármacos, em casos de erros médicos ou invalidez (Dean, Victor, Boone, Philpott, & Hess, 2009). O objectivo do presente trabalho é validar um protocolo de TVS na população idosa, nomeadamente numa amostra da comunidade, num grupo com Declínio Cognitivo Ligeiro e numa amostra instruída a "simular".

I – Enquadramento conceptual

1. Definição de simulação

Têm sido propostas definições e critérios para “simulação”, entre os quais a do DSM-IV-TR que define “simulação” como a “*produção intencional de sintomas físicos ou psicológicos falsos ou exagerados, motivados por incentivos externos*” (DSM-IV-TR; APA, 2002, p. 739). São referidas algumas situações que podem indicar a maior probabilidade de ocorrência destes comportamentos como: contexto médico-legal; presença de perturbação antisocial da personalidade; discrepância entre resultados objectivos da avaliação e incapacidade referida pelo sujeito; e falta de colaboração/adesão ao tratamento. Realça-se nesta definição três aspectos: a) invenção/exagero de sintomas; b) motivação ou intenção consciente do sujeito para simular; e c) presença de incentivos externos.

Têm sido enumeradas algumas limitações à definição proposta pelo DSM-IV-TR, nomeadamente que a “simulação” é mais que o exagero de sintomas, que deve ser intencional, e identificada a presença de incentivos externos (Rogers & Ganacher, 2011). Estes autores referem ainda que todas as avaliações forenses envolvem um contexto médico-legal, pelo que este aspecto não deve ser usado pelos profissionais na distinção entre respostas de “simulação” ou “honestas”. Por outro lado, alguns autores (Larrabee, 2005, cit. in Simões et al., 2010) têm realçado a necessidade dos comportamentos de simulação serem considerados num *continuum*, sendo difícil distinguir entre sintomatologia intencional ou não intencional.

A dicotomia destes critérios (incentivos externos *versus* internos, comportamentos volitivos *versus* inconscientes) em contexto prático tornam-se difíceis de distinguir, além de poderem coexistir, por exemplo, incentivos internos e externos. Foi com base nestas limitações que alguns autores propuseram definições e critérios alternativos.

Rogers (1990, cit. in Slick, Sherman, & Iverson, 1999) propôs critérios de diagnóstico específicos para a simulação de perturbações psiquiátricas que abarcam a recolha de informação de diferentes domínios, entre os quais auto relatos, pontuações nos testes, observação do comportamento e informação colateral. Usando um modelo similar ao DSM-IV, especificou um número mínimo de critérios em cada domínio necessários para o diagnóstico de simulação (ex. número atípico de sintomas raros, informação colateral contraditória, evidência de exagero ou fabrico de

sintomas em testes estandardizados), identificando também critérios de exclusão (presença de perturbação factícia). Rogers (1990) ainda sugeriu três modelos explicativos de simulação: o Patogénico; Criminológico e Adaptativo. O Modelo Patogénico aplica-se a pessoas que estão motivadas a simular por patologia subjacente. O Modelo Criminológico postula que a simulação ocorre por maldade em que os sujeitos irão fingir perturbações mentais para obter recompensas. Por seu turno, o Modelo Adaptativo postula que a simulação é uma tentativa, pelo menos na perspectiva do sujeito, de ser bem-sucedido em circunstâncias adversas (Rogers, Sewell, & Goldstein, 1994), em que este decide simular com base na probabilidade e utilidade esperada de tal comportamento (Slick et al., 1999). Já Greiffenstein e colaboradores (1994, cit. in Slick et al., 1999) propuseram um conjunto de critérios de diagnóstico de simulação de défices mnésicos, para uso em contexto neuropsicológico, a saber: a) desempenho pobre em 2 ou mais medidas neuropsicológicas; b) incapacidade total na maioria dos papéis sociais; c) contradições entre informações colaterais e os relatos; d) perda de memória remota. Contudo, Slick e colaboradores (1999) apontam algumas críticas, entre as quais: o facto de não apresentarem uma definição de simulação; não especificação de condições de exclusão ou de diagnóstico diferencial; não inclusão de observações comportamentais; não especificação de alguns dos critérios; ser restrito a défices de memória.

Slick e colaboradores (1999) desenvolveram assim um conjunto de critérios de diagnóstico de Simulação de Perturbação Neurocognitiva (SPN) onde a simulação é definida como “*exagero ou produção intencional de uma perturbação cognitiva realizado com o propósito de obter ganhos materiais substanciais, evitar ou escapar a responsabilidades formais ou de trabalho*” (Slick et al., 1999; p.552). Os critérios de diagnóstico de SPN são: a) presença de incentivos externos; b) evidências na avaliação neuropsicológica; c) evidências de exagero nos auto-relatos do sujeito; d) assegurar que estes comportamentos não podem ser melhor explicados por doença psiquiátrica, neurológica ou perturbação desenvolvimental. Os autores propõem o diagnóstico de SPN definitiva, provável ou possível consoante são preenchidos determinados critérios (cf. Anexo 1).

Por outro lado, Ferguson (2004, cit. in Simões, 2006) sugere a existência de quatro situações distintas de simulação: a) simulação pura ou

falsificação de défices, em que os sujeitos não apresentam défice decorrentes de uma lesão mas procuram alegar a sua presença (forma rara de “simulação”); b) simulação parcial ou exagero de sintomas, em que o sujeito manifesta dificuldades genuínas mas maximiza-as (forma mais comum de “simulação”); c) manutenção ou agravamento de sintomas, no qual o sujeito já apresentou previamente défices mas, após recuperação ou melhoria, relata que os sintomas persistem ou agravam-se; d) atribuição incorrecta dos sintomas ou falsa imputação, em que o sujeito apresenta sintomas não relacionados com a lesão mas atribui-os intencionalmente à mesma.

O que também pode ocorrer aquando das avaliações e justificar desempenhos mais baixos é o esforço insuficiente. Assim, em alguns contextos, faz mais sentido pensar em esforço insuficiente, caracterizado pelo “*desempenho alcançado num teste que é significativamente mais baixo do que os padrões de desempenho conhecidos associados a perturbações neurológicas verdadeiras*” (Simões, 2005, p. 456). A utilização deste conceito torna-se mais compreensiva, até pelos múltiplos factores (ex. motivacionais, situacionais/contextuais) que influenciam os desempenhos e, consequentemente, os resultados nas avaliações realizadas.

Têm sido propostos diferentes termos na literatura para as situações em que os sujeitos obtêm pontuações abaixo dos pontos de corte definidos para cada medida, nomeadamente esforço insuficiente, esforço reduzido, esforço inválido, *faked effort*, desempenho não credível, simulação. Bigler (2012) e Larrabee (2012) sugerem o abandono do termo “esforço” em contextos neuropsicológicos (em contextos forenses estes termos podem ser adequados), arguindo que se os TVS são construídos de forma a que seja improvável sujeitos mesmo com défices reais errarem, são pouco exigentes do ponto de vista cognitivo. Nas avaliações neuropsicológicas os TVS são utilizados para inferir a validade das inferências, pelo que a utilização de termos como validade de desempenho (*performance validity*) e validade de sintomas (*symptom validity*) é mais apropriada. Larrabee (2007) defende que os TVS são medidas de validade de desempenho visto clarificarem, nestes contextos, em que medida o desempenho dos sujeitos nos testes é ou não o reflexo das suas actuais capacidades.

A possibilidade de simulação ou exagero de sintomas deve ser objectivamente examinada pelos profissionais quando: o paciente se

encontra em contexto medicolegal; estão presentes incentivos para exagero de sintomas; os sintomas não têm sentido médico, neuropsicológico ou psicológico; a queixa de incapacidade é excessiva comparativamente a resultados objectivos; falta de cooperação durante a avaliação; inconsistências entre as queixas e os comportamentos observados; ou contradições entre os auto-relatos e os registos médicos presentes. Assim, a avaliação da simulação ou exagero de sintomas é fundamental em muitos casos clínicos, medicolegais e forenses, isto porque os resultados obtidos nestes contextos determinam a recepção de ganhos financeiros ou pessoais, ou mesmo o início de determinado tratamento (Teichner & Wagner, 2004), não descurando que a administração de TVS deve ser distribuída ao longo da avaliação, integrando-os em baterias mais alargadas (Simões et al., 2010).

2. Abordagens na detecção de simulação

Até aos anos 80, a detecção de simulação, exagero de sintomas e/ou esforço insuficiente baseava-se na análise qualitativa das respostas a testes mais clássicos como WAIS, WMS, *Trail Making Test*, MMPI-2, Figura Complexa de Rey (Tombaugh, 1997), sendo os padrões de respostas inconsistentes o principal indicador de simulação. Nestes casos, as situações mais comuns que podem sugerir a presença destes padrões de resposta incluem os erros ou respostas aproximadas, decréscimos pronunciados na evocação tardia, discrepâncias entre pontuações nos testes que medem os mesmos processos, erros em itens mais fáceis e acertos nos mais difíceis, e inconsistências entre queixas observadas e dados observacionais do comportamento/registos clínicos (Tombaugh, 1997).

Contudo, estes testes nem sempre identificam correctamente os casos de simulação, exagero de sintomas ou esforço insuficiente, o que incentivou alguns autores a desenvolver testes especificamente construídos para a avaliação da validade de sintomas, os testes de Validade de Sintomas (TVS) (cf. Tombaugh, 1997; Franklin, 2008). Constituídos por um elevado número de itens os TVS são usados especificamente para detecção de simulação e/ou esforço insuficiente. São testes que normalmente usam o paradigma de escolha forçada e, idealmente, são construídos de modo a detectar simulação de défices mas insensíveis a défices reais e ao impacto de variáveis socio-demográficas (Tombaugh, 1997; Ashendorf et al., 2004). Os TVS baseiam-

se no princípio “efeito de chão”, onde os itens são de tal forma simples que podem ser respondidos mesmo por sujeitos com défices cognitivos reais. Os simuladores tendem a responder erradamente a estas tarefas bastante simples pois acreditam que sujeitos com défices reais responderiam incorrectamente (Franklin, 2008).

O desenvolvimento e validação de diferentes medidas de simulação ou esforço insuficiente é essencial por diversos motivos. Primeiramente, o uso de diferentes medidas permite a identificação convergente de simulação ou exagero de sintomas e resulta numa conclusão mais fidedigna do que a baseada no recurso a apenas uma medida. Além do mais, o recurso a diferentes testes ajuda no processo de detecção de simulação (é menos provável que um sujeito possa aprender a fingir se forem administradas no protocolo de avaliação diversas medidas). Por fim, os simuladores também são selectivos nos domínios que “escolhem” simular (memória, capacidade de leitura, etc.) pelo que o recurso a diferentes medidas permite melhor “captar” todos estes domínios (Boone et al., 2000). Contudo, com a proliferação dos TVS e incremento de investigação na área, emerge não só a problemática de saber que testes aplicar mas, também, a possibilidade da obtenção de dados contraditórios que poderão advir das diferentes medidas aplicadas, o que desencoraja muitas vezes os profissionais no uso dos TVS (Rosenfeld, Green, Pivovarova, & Dole, 2010; Bliger, 2012). A utilidade do uso de diferentes medidas de simulação ou esforço insuficiente está bem documentada na literatura (cf. Boone et al., 2000; Simões et al., 2010; Rosenfeld et al., 2010). Contudo, os estudos que referem o uso de diferentes TVS não analisam o contributo de cada medida. Atendendo a este facto, Rosenfeld e colaboradores (2010) verificaram que apesar da possibilidade de ocorrência de resultados contraditórios, o uso de medidas adicionais aumenta a capacidade de detecção de “simulação” (Rosenfeld et al., 2010).

Apesar do suporte empírico e da sua relevância, têm sido apontadas algumas críticas à investigação com os diferentes TVS, nomeadamente que:

- i) os valores de especificidade¹ e sensibilidade adequados são obtidos

¹ *Sensibilidade* diz respeito à capacidade de um teste identificar correctamente uma característica quando esta está presente (neste caso a simulação). Por outro lado, a *especificidade* refere-se à capacidade de determinado teste excluir correctamente aqueles que não possuem determinada característica. Enquanto um elevado valor de sensibilidade está associado a um valor reduzido de falsos negativos, quanto maior o valor de especificidade menor a taxa de falsos positivos.

sobretudo em planos de investigação com desenhos experimentais (com sujeitos instruídos a responder como se tivessem motivos para simulação) (Clegg, Fremouw, & Mogge, 2009; Greve, Etherton, Ord, Bianchini, & Curtis, 2009; Rosenfeld et al., 2010), tanto mais que um sujeito com um défice real pode também simular ou exagerar défices (Franklin, 2008; Yochim, Kane, Horning & Pepin, 2010) na medida em que a gravidade destes vai condicionar a recompensa; ii) a maioria das medidas focam-se num único domínio de funcionamento; iii) escassez de estudos do desempenho de diferentes grupos e com base em diferentes medidas; iv) taxas de falsos negativos verificadas com cada medida singularmente.

A validação de um protocolo de avaliação nestes contextos torna-se assim indispensável, como solução para as taxas de falsos negativos encontradas separadamente com cada medida. Além do mais uma pontuação isolada num TVS não é suficiente para concluir a presença de respostas enviesadas (Lee, Loring, & Martin, 1992; Tombaugh, 1997; Boone et al., 2000; Lee et al., 2000; Clegg et al., 2009; Simões et al., 2010; Rudman et al., 2011) pelo que é necessário o uso convergente de vários métodos e medidas.

3. Testes de validade de sintomas usados na detecção de simulação ou esforço insuficiente de défices cognitivos

3.1 *Test of Memory Malinger* (TOMM)

O TOMM é um teste de reconhecimento visual originalmente desenvolvido por Tombaugh, em 1996, para adultos. Tombaugh (1997) num estudo de validação realizou 4 experiências (com amostras cognitivamente intactas e com défice neurológico – TCE, afasia, défice cognitivo e demência), tendo verificado que os resultados neste instrumento são relativamente insensíveis a défices de memória bem como à idade e escolaridade e promissor na detecção de “simulação” ou esforço insuficiente. Um resultado <45 no Ensaio 2 e Ensaio de Reconhecimento classifica correctamente 91% do total dos participantes, 95% dos sujeitos sem demência, e 100% dos controlos. Verificou ainda que os testes de memória de reconhecimento de imagens representam um paradigma em que o grau de dificuldade percebido da tarefa excede a dificuldade real da mesma. Estes dados foram corroborados por Rees, Tombaugh, Gansler e Moczynski (1998) que também verificaram que o TOMM é capaz de detectar os casos

de simulação/exagero de sintomas de défices mnésicos.

Noutra pesquisa Teichner e Wagner (2004) examinaram o desempenho no TOMM numa amostra de idosos que incluía casos de demência e casos sem e com défice cognitivo (ex. doença de Parkinson, álcool). Os sujeitos cognitivamente intactos e com défice cognitivo manifestaram desempenhos similares no TOMM, com o ponto de corte <45 no Ensaio 2 e Ensaio de Reconhecimento. Já o grupo de sujeitos com diagnóstico de demência apresentou desempenhos inferiores que os outros dois grupos em todos os ensaios, verificando-se que mesmo pontos de corte menos robustos não classificaram correctamente os sujeitos como “honestos”. Este estudo, além da inclusão de sujeitos idosos, demonstra a especificidade do TOMM e a sua utilidade, refutando a sua insensibilidade a estados demenciais.

Fernandes (2009) procurou analisar o perfil dos resultados no TOMM em adultos idosos (Controlo e Declínio Cognitivo Ligeiro). Verificou que este parece ser insensível a variáveis sociodemográficas, afectivas e capacidade funcional. Estes resultados sugerem a utilidade do instrumento na avaliação da simulação ou exagero de sintomas na população idosa, mesmo com declínio cognitivo, e com os pontos de corte propostos inicialmente pelo autor do teste.

3.1.1 Versões abreviadas do TOMM

O facto dos TVS não serem incluídos em baterias de avaliação neuropsicológica pode dever-se a diversos factores, entre os quais: falta de conhecimento por parte de alguns profissionais acerca da existência; bases teóricas e de formação distintas (Bauer et al., 2007); bem como o tempo de administração que estes instrumentos exigem (O’Byrant et al., 2007; Bauer et al., 2007; O’Byrant et al., 2008). A identificação duma medida breve de rastreio de “simulação” ou de esforço insuficiente poderá aumentar o recurso dos profissionais aos TVS em contextos clínicos e de investigação. Foi neste plano que alguns autores (O’Byrant et al., 2007; Bauer et al., 2007; O’Byrant et al., 2008) procuraram perceber se existia evidência que justificasse a administração do TOMM circunscrita apenas ao Ensaio 1. Os resultados encontrados por O’Byrant e colaboradores (2007; 2008) corroboraram a hipótese de que o Ensaio 1 é uma opção viável como medida de *screening* de simulação ou esforço insuficiente, dados corroborados por Bauer e

colaboradores (2007). O'Bryant e colaboradores (2008) e Fernandes (2009) demonstraram ainda que 100% dos sujeitos que obtiveram pontuações >45 no Ensaio 1 continuaram com esse desempenho nos seguintes, corroborando a possibilidade de terminar a administração da prova neste ensaio. Os dados obtidos no estudo de Fernandes (2009) apontam para um ponto de corte ≥ 33 , no Ensaio 1, como traduzindo 100% de probabilidade dos sujeitos serem identificados no TOMM como “não simuladores”.

É contudo ressalvado que estes desempenhos não devem ser considerados como evidência definitiva de ausência ou presença de simulação, ressalvando que quando existe sugestão de simulação devem ser implementadas avaliações mais exaustivas para uma maior precisão diagnóstica (O'Bryant et al., 2007; O'Bryant et al., 2008; Bauer et al., 2007).

Na sequência destas conclusões relativamente à aparente baixa utilidade do Ensaio 2 e Ensaio de Reconhecimento quando as pontuações no Ensaio 1 são superiores a 45, Greve e Bianchini (2006) procuraram verificar se o Ensaio de Reconhecimento oferece alguma validade adicional. Concluíram que, através da análise do desempenho de uma amostra de sujeitos com TCE e pacientes com dor crónica (ambos com presença de incentivos externos), a não administração do Ensaio de Reconhecimento resulta em 3% de falsos negativos. O custo da não aplicação deste ensaio foi minimizado quando se recorreu a pelo menos um outro TVS.

Hilsabeck, Gordon, Hietpas-Wilson, & Zartman (2011) sugerem um ponto de corte ≥ 41 como adequado para detectar esforço suficiente, e um ponto de corte ≤ 25 como indicativo de ausência de validade do desempenho, pelo que resultados superiores ou inferiores a estes pontos de corte, respectivamente, tornam desnecessária a aplicação dos restantes ensaios, em contextos clínicos.

Denning (2012), além da pontuação no Ensaio 1, procurou verificar também a eficácia do número de erros nos 10 primeiros itens. Verificou que, numa amostra clínica de 497 pacientes com idades compreendidas entre os 21 e 87 anos (excepto pacientes com diagnóstico de demência), os valores de sensibilidade para o Ensaio 1 (ponto de corte ≤ 40) e para os erros nos 10 primeiros itens (ponto de corte ≥ 1) foram de 72 e 71%, respectivamente. Os valores de especificidade foram de 87% e 86%, respectivamente, o que corrobora a capacidade destes índices na detecção de simulação ou esforço

insuficiente.

Existe assim evidência do TOMM ser uma medida útil na detecção de “simulação” ou exagero de sintomas, excluindo casos de sujeitos com diagnóstico de demência. Uma versão abreviada deste instrumento pode constituir uma forma de *screening* de avaliação da motivação e desempenho dos sujeitos.

3.2 Dot Counting Test (DCT)

O DCT, originalmente desenvolvido em 1941 por André Rey, consiste em 2 grupos de cartões, em que estão apresentados conjuntos de pontos agrupados ou distribuídos aleatoriamente. Supõe-se que, em casos de simulação ou esforço insuficiente, o tempo requerido para contar os pontos agrupados é superior ao tempo para contar os não agrupados (Boone et al., 2002a). Boone e colaboradores (2002a) analisaram o desempenho de sujeitos com suspeita de “simulação” (incapacidade/dano pessoal e prisão hospitalar), e de 9 grupos clínicos (lesão cerebral, dificuldades aprendizagem, AVC, esquizofrenia, idosos normais, idosos deprimidos, demência ligeira e moderada) nas diferentes medidas deste teste. Pontos de corte de >7 segundos para o tempo médio dos pontos agrupados, > 3 erros, valor da razão entre média do tempo dos pontos não agrupados pelos agrupados ≤ 1.5 e resultado combinado (soma de média de tempo para contar pontos agrupados, não agrupados e total de erros) ≥ 17 tiveram altos valores de especificidade ($\geq 85\%$) para as amostras clínicas combinadas (não incluída demência moderada). Com o ponto de corte ≥ 17 para o resultado combinado verifica-se uma sensibilidade de 100% e 75% no grupo de prisão hospitalar e de incapacidade, respectivamente, enquanto este valor é superior a 90% para todos os grupos clínicos combinados, com exceção da demência moderada. Verificou-se ainda que os dois grupos de suspeita de simulação comportaram-se de diferentes formas consoante o contexto onde se inserem (possivelmente por estarem à procura de recompensas diferentes), o que demonstra que os pontos de corte obtidos com um subgrupo não se generalizam necessariamente para outro.

Weiss e Rosenfeld (2010) procuraram analisar a utilidade do DCT numa amostra rural da região da Índia partindo do pressuposto que a capacidade exigida para a sua realização – contar – é comum em todas as

culturas. Os resultados obtidos indicam que uma grande proporção de sujeitos foi classificada como “simuladores”, sendo que apenas uma pequena quantidade foi considerada pelos profissionais que os avaliaram como estando realmente a exagerar sintomas. Uma explicação apontada para estes resultados reside na natureza da amostra (rural e com níveis de escolaridade baixos), distinta da amostra americana de validação do DCT. Este estudo realça que os pontos de corte obtidos para uma cultura nem sempre se generalizam a outras, podendo assim os sujeitos serem erradamente considerados como “simuladores”. Verificaram ainda que a depressão e perturbações psiquiátricas têm pouco impacto no desempenho neste teste.

3.3 Rey 15-Item Memory Test (15-IMT)

O 15-IMT, criado por Rey em 1964, é um teste de memória visual para detectar simulação ou exagero de sintomas de queixas de memória. É apresentado como um teste em que os sujeitos têm de memorizar e evocar 15 itens, apesar de na realidade apenas requerer a memorização de cinco unidades (Lee et al., 1992). Estes autores administraram o teste a sujeitos internados com epilepsia do lobo temporal (défices de memória genuínos) e, ainda, a pacientes de ambulatório que tinham sido encaminhados para avaliação neuropsicológica e verificaram que os pacientes de ambulatório em situação de litígio tiveram pior desempenho que os com epilepsia e pacientes de ambulatório em situação de não litígio, sendo que estes últimos não diferiram entre si. Os resultados sugerem que, com um ponto de corte <7, o teste é útil na detecção de possível simulação de queixas de memória.

Posteriormente, o 15-IMT foi objecto de um aperfeiçoamento, com o intuito de melhorar os seus valores de sensibilidade e especificidade, introduzindo-se ao formato inicial (Evocação Imediata) um Ensaio de Reconhecimento (Boone, Salazar, Lu, Warner-Chacon & Razani, 2002b). Estes autores encontraram altos valores de especificidade (97%-100%) para um ponto de corte <9 no Ensaio de Evocação Imediata, e valores de sensibilidade razoáveis (47%). A administração do Ensaio de Reconhecimento aumenta os valores de sensibilidade (71%), mantendo elevados valores de especificidade ($\geq 92\%$), corroborando a sua utilidade.

Numa meta análise, Reznick (2004) concluiu que este é um instrumento com valores de especificidade adequados, mas com taxas de

falsos negativos elevadas, evidenciando a necessidade de ajuste dos pontos de corte. Têm sido propostos diferentes pontos de corte para o Ensaio de Evocação Imediata deste teste. Simões e colaboradores (2010) realizaram um estudo com uma amostra de adultos idosos (saudáveis, com problemas cognitivos e depressão) com o intuito de analisar o desempenho comparativamente com uma amostra de controlo (adultos jovens). Os dados obtidos evidenciam elevados valores de especificidade para o ponto de corte <9 e <20 no ensaio de Evocação Imediata e no Resultado Combinado², respectivamente, mas apenas no grupo de jovens adultos. Concluíram os autores a necessidade dos pontos de corte serem ajustados para a população idosa, sugerindo para o grupo de adultos idosos saudáveis e com depressão um ponto de corte no Ensaio de Evocação Imediata para valores próximos de 6, sendo que este ponto de corte deve ser ainda mais reduzido para o grupo com problemas cognitivos (Declínio Cognitivo Ligeiro e Demência Ligeira).

As vantagens do 15-IMT incluem o recurso a poucos materiais de teste, administração breve e o sistema de cotação simples. Tais vantagens contribuem para que esta medida seja bastante usada e estudada para avaliação da simulação e/ou esforço insuficiente. Devido à necessidade inquestionável de dados normativos para as diferentes populações e culturas, Strutt, Scott, Shrestha e York (2011) procuraram examinar a influência de variáveis socio demográficas neste instrumento e fornecer dados normativos por idade e escolaridade para a comunidade espanhola residente nos Estados Unidos. Numa amostra composta por 130 adultos idosos (50-69 anos) falantes espanhóis cognitivamente intactos Strutt e colaboradores verificaram que estes sujeitos tinham pontuações mais baixas que os sujeitos ingleses. Os pontos de corte definidos para a população inglesa não se revelaram recomendáveis para utilização com população espanhola residente nos Estados Unidos, reforçando a necessidade dos resultados nas diferentes medidas deverem ser interpretados com cautela, quando não existem normas para a população em causa.

3.4 b-test

O *b-test* foi construído e desenvolvido tendo subjacente o pressuposto

² Número de itens correctamente evocados + (número de itens correctamente reconhecidos – número de falsos positivos).

de que os sujeitos têm crenças erradas acerca das consequências neuropsicológicas da lesão cerebral, assumindo que tais lesões provocam défices na memória de reconhecimento, capacidade de atenção, informação previamente aprendida, força motora e destreza. A tarefa dos sujeitos neste teste consiste em rodear todas as letras “b” num conjunto de 15 páginas. As medidas possíveis deste teste incluem: o número de erros (circular letras “d”); número de erros de comissão (circular figuras que não “b”); número de erros de omissão (não circular letra “b”) e o tempo total para completar a tarefa (Boone et al., 2000). Os autores usaram o *b*-test numa amostra de simuladores (em situação de litígio ou de procura de obtenção/manutenção de recompensa por incapacidade) e em vários grupos clínicos (lesão cerebral moderada a severa, idosos deprimidos, dificuldades de aprendizagem, esquizofrenia, AVC, idosos normais), verificando que os sujeitos suspeitos de simulação tinham piores desempenhos que os restantes grupos. Quando usado um ponto de corte de >2 erros de comissão, 76,5% dos suspeitos de simulação foram correctamente identificados, com valores de especificidade de 100% e 82.6% para os sujeitos com lesão cerebral e todos os grupos clínicos combinados, respectivamente. Relativamente aos erros de omissão e ao tempo para completar a tarefa obtiveram-se valores de sensibilidade baixos, mas a especificidade continuou elevada. Um ponto de corte >40 para os erros de omissão classificou correctamente mais de metade dos sujeitos suspeitos de simulação (58.8%) com valores de especificidade 95% para os sujeitos com lesão cerebral e 85.1% para os grupos combinados; um ponto de >12 minutos para o tempo também classifica correctamente mais de metade dos simuladores (57.6%), com valores de especificidade na ordem dos 85% para os sujeitos com lesão cerebral e 83.9% para todos os grupos comparados combinados. Apesar de se verificar que o tempo total de resposta está positivamente correlacionado com a idade, o mesmo não se verifica em relação aos erros de omissão e comissão.

4. Testes de Validade de Sintomas usados na detecção de simulação ou esforço insuficiente de défices psicopatológicos

4.1 Structured Inventory of Malingered Symptomatology

Dos instrumentos de auto-relato construídos para detecção do exagero de sintomas psicopatológicos, o *Structured Inventory of Reported Symptoms* (SIRS; Rogers et al., 1992) é a medida mais extensivamente validada (Jelicic, Ceunen, Peters, & Merckelbach, 2011) e consiste numa entrevista estruturada. Existem medidas mais breves, entre os quais o *Structured Inventory of Malingered Symptomatology* (SIMS) (Clegg et al., 2009), referenciado na literatura como capaz de avaliar diferentes domínios (psiquiátrico, cognitivo, e físico) (Wisdom, Callahan, & Shaw, 2010). Clegg e colaboradores (2009) concluíram que, usando os pontos de corte propostos de >14 ou >16 para o SIMS e >13 na escala de simulação do ADI (*Assessment of Depression Inventory*), se obteria bastantes falsos positivos no SIMS, e bastantes falsos negativos no ADI. Posto isto, sugerem um aumento do ponto de corte para a pontuação total do SIMS >19 e um ponto de corte >9 no ADI. Neste estudo as subescalas do SIMS são menos eficazes que a pontuação total na diferenciação de simuladores ou não simuladores (Clegg et al., 2009). Wisdom e colaboradores (2010) verificaram que os sujeitos “simuladores” tiveram desempenhos piores em todas as medidas de simulação comparativamente aos “honestos”. Com um ponto de corte de 24 não existiram falsos positivos, apesar de reduzir a capacidade para detectar verdadeiros “simuladores”. Contudo, é realçada a maior utilidade de pontos de corte que favorecem a especificidade em prol da sensibilidade, evitando assim falsas acusações de simulação.

Sabendo que os sujeitos podem ser “treinados” antes das avaliações neuropsicológicas (recurso a informação da Internet, indicações fornecidas pelos advogados) acerca dos desempenhos esperados nos diferentes TVS (Jasinski et al., 2011) Jelicic e colaboradores (2011) procuraram investigar o efeito de diferentes tipos de *coaching* no TOMM e no SIMS. Os autores colocaram a hipótese de instruções (*coaching*) mais longas (isto é, não informar apenas acerca dos sintomas envolvidos mas também acerca do teste), influenciarem a sensibilidade diagnóstica destes instrumentos. Os

resultados evidenciam que tanto o TOMM como o SIMS são relativamente resistentes aos efeitos do *coaching*.

5. Outras medidas usadas na detecção de simulação e/ou esforço insuficiente

5.1 Subteste Memória de Dígitos e Memória Lógica

As escalas de inteligência e memória de Wechsler são bastante usadas na prática clínica dos neuropsicólogos (Lu, Rogers, & Boone, 2007; Jasinski et al., 2011), o que torna viável a detecção de índices dos diferentes subtestes para detecção de possível esforço insuficiente/simulação. Estas alternativas são úteis uma vez que não implicam a necessidade de tempo extra de administração nem aquisição de materiais adicionais, reduzem a possibilidade dos sujeitos perceberem que estão a ser avaliados com este objectivo, bem como torna menos provável a possibilidade de serem ensinados acerca do desempenho desejável tal como nos TVS (Swihart, Harris, & Hatcher, 2008; Jasinski et al., 2011).

A meta análise de Jasinski e colaboradores (2011) mostra a utilidade do subteste Memória de Dígitos (MD) na detecção de simulação de défice neurocognitivo. A MD faz parte tanto da Escala de Inteligência de Wechsler para adultos (WAIS-R, WAIS-III, WAIS-IV) como da Escala de Memória de Wechsler (WMS-R, WMS-III, WMS-IV), é de rápida aplicação, e não sofreu alterações importantes nas diferentes revisões. A meta análise englobou 24 estudos, onde analisaram especificamente a capacidade de dois índices deste subteste na diferenciação entre “simuladores” e “honestos”, nomeadamente a Fiabilidade da Memória de Dígitos (*Reliable Digit Span*, correspondente à soma da sequência mais longa de dígitos repetida sem erros na condição sentido directo e inverso) e Memória de Dígitos – pontuação dos resultados de acordo com idade (*Digit Span Age-Corrected Scale Score*). Verificaram que a Fiabilidade da Memória de Dígitos (FMD) obteve valores de especificidade e sensibilidade de 86.1% e 63.3%, respectivamente. No índice Memória de Dígitos – pontuação dos resultados de acordo com idade – os valores são de 86.5% e 59.7%, respectivamente. Não se verificaram diferenças significativas entre os dois índices, o que sugere que os profissionais possam escolher entre os dois, apesar do segundo utilizar resultados estandardizados, podendo se tornar mais difícil para os

simuladores (ou pessoas que exageram sintomas) avaliar a quantidade de “simulação requerida” para produzir um resultado favorável (Jasinski et al., 2011). Estudos prévios demonstraram que a FMD (≤ 6) e o resultado padronizado da MD da WAIS-III (≤ 5) podem ser usados como indicadores de simulação também em casos de alegada exposição a tóxicos (Greve et al., 2007).

Killgore e DellaPietra (2000, cit. in Lu et al., 2007) realizaram um estudo com o subteste Memória Lógica da Escala de Memória de Wechsler (WMS) no qual identificaram seis itens do Ensaio de Reconhecimento (Índice de Respostas Infrequentes - *Rarely Missed Index*, RMI), nos quais mesmo os indivíduos que não conhecem as histórias dão respostas correctas. Este índice é calculado através da soma da pontuação ponderada atribuída a cada um dos itens (item 12, 16, 18, 22, 24, 29) (cf. Anexo 2), sendo que pontuações ≤ 136 são indicativas de esforço insuficiente (Killgore & DellaPietra, 2000, cit. in Lu et al., 2007; Swihart et al., 2008). Este ponto de corte classificou correctamente 97% dos 36 “simuladores” (instruídos a simular lesão cerebral), estando também associado a elevados valores de especificidade (100%), distinguindo deste modo entre simuladores e pacientes com défice neurológico (Killgore & DellaPietra, 2000, cit. in Lu et al., 2007). Swihart e colaboradores (2008) realizaram um estudo de replicação dos estudos iniciais com o RMI e verificaram que, no uso clínico, o índice RMI classificou incorrectamente sujeitos “honestos” e que não detectou correctamente todos os sujeitos que estavam a “simular”. Concluíram que, em condições mais realísticas e similares ao contexto da prática clínica³, o RMI não é ideal para detectar “simuladores”.

6. Condições Clínicas: Demência, Depressão e ansiedade

A investigação que incida no desempenho da população com demência, maioritariamente idosa, em provas de simulação ou esforço insuficiente é relativamente escassa. Para que a interpretação dos resultados nos diferentes TVS seja eficaz, é importante conhecer quais os testes que

³ No estudo inicial na instrução era pedido que os sujeitos "simulassem défices de memória" e foi apenas aplicado o subteste Memória Lógica. Neste estudo, os sujeitos eram instruídos a simular dificuldades cognitivas devido a uma lesão cerebral, foram administrados diversos testes e, conseqüentemente, existem maiores possibilidades de escolherem onde simular (ex. tarefas de memória verbal, visual ou tarefas de outra natureza).

proporcionam menor taxa de falsos positivos, de perceber a relação entre a severidade da demência e os valores de especificidade nos diferentes testes, bem como saber se os pontos de corte tradicionais necessitam ou não de ser ajustados a esta população. Com base nestes pressupostos, Dean e colaboradores (2009) procuraram analisar e examinar o desempenho numa amostra com demência em diferentes índices (MD – pontuação dos resultados de acordo com idade, Índice FMD, MD - tarefa cronometrada (3 e 4 dígitos sentido directo), Vocabulário menos Memória de Dígitos, DCT - total erros, TOMM - Ensaio 2, Teste de Memória de Reconhecimento de Warrington – Palavras, 15-IMT, WMS-III - RMI, *Finger Tapping* - mão dominante, *b-Test* - total erros, *Rey Word Recognition Test*, *Rey Auditory Verbal Learning Test* - equação de esforço, *Rey-Osterreith* - equação de esforço, *Rey-Osterreith/RAVLT*), derivados de vários testes. Apenas o Índice Vocabulário menos Memória de Dígitos apresentou valores de especificidade >90% para os casos de demência, independentemente da sua severidade. Os resultados deste estudo demonstraram que: (i) quanto maior o grau de severidade da demência, maior a probabilidade dos sujeitos serem erradamente classificados como “simuladores”; (ii) a maior parte TVS analisados apresentam uma taxa elevada de falsos positivos; e (iii) ajustar os valores de especificidade para que as medidas tenham valores de especificidade superiores a 90% implica muitas vezes o custo de diminuição da sensibilidade do mesmo. Relativamente a este último ponto refere-se o exemplo do Ensaio de Evocação Imediata do 15-IMT. É sugerido um ponto de corte <1 tornando-se evidente que os sujeitos não teriam de evocar nenhum item para não serem considerados “simuladores”. Neste estudo, surgem como medidas promissoras para usar com esta população o Vocabulário menos Memória de Dígitos, e 3 e 4 dígitos cronometrados em sentido directo.

Rudman, Oyebode, Jones, e Bentham (2011) realizaram um estudo com sujeitos em idade activa diagnosticados com demência. Como hipóteses específicas consideraram que: a) funcionamento cognitivo estaria relacionado com o desempenho em cada teste; b) haveria diferença no desempenho em cada teste de acordo com o grau de severidade da demência; c) existiriam diferenças nas diferentes taxas de acertos e erros entre os diferentes testes; e d) os desempenhos nos TVS seriam influenciados por

domínios específicos do funcionamento cognitivo. Para corroborar/refutar estas hipóteses os autores aplicaram 6 TVS (entre os quais o TOMM, DCT e 15-IMT) e medidas de avaliação do estado emocional. Tal como hipotetizado verificaram que os défices cognitivos estão correlacionados com um desempenho mais baixo nos diferentes testes. Em relação à segunda hipótese os sujeitos com demência ligeira tiveram melhores resultados que aqueles com demência moderada/severa, sendo o 15-IMT e o TOMM os testes mais sensíveis à severidade da disfunção cognitiva. Foi possível corroborar também que os sujeitos com demência moderada/severa obtiveram mais erros que os com demência ligeira. Desta investigação ressalva-se ainda que apesar do TOMM ser um teste de reconhecimento visual, os resultados fornecem evidência de codificação multimodal e que os processos de produção verbal contribuem para a codificação de estímulos visuais, o que justifica que participantes com dificuldades de nomeação e semânticas devido à demência tenham pior desempenho no TOMM. Ainda foi possível concluir que o 15-IMT e o TOMM são medidas sensíveis a défices de memória e correlacionadas com o funcionamento cognitivo, congruente com estudos prévios (cf. Teichner & Wagner, 2004; Simões et al., 2010). Relativamente ao estado emocional verificou-se que os resultados corroboram igualmente estudos prévios (cf. Ashendorf et al., 2004; Lee et al., 2000) que reportam que a ansiedade e depressão não têm influência no desempenho nos diferentes TVS. O DCT é neste estudo o único com evidência para ser usado em sujeitos com possível quadro demencial.

No que concerne aos estados demenciais, existe evidência que são poucas as medidas que produzem valores de especificidade adequados e consequentemente baixas taxas de falsos positivos. A maioria das medidas de simulação ou esforço insuficiente revelam-se sensíveis e pouco adequadas a esta população, com base nos referenciais propostos para as restantes populações (cf. Dean et al., 2009; Simões et al., 2010; Rudman et al., 2011). Estar consciente de que défices cognitivos desta ordem influenciam os resultados nos diferentes testes é essencial, sob pena de considerar erradamente como simuladores sujeitos com défices reais (com diagnóstico de demência).

Relativamente às alterações emocionais, a possibilidade da depressão e ansiedade influenciar o desempenho nos testes coloca a hipótese dos dados

obtidos poderem ser melhor explicados por estas variáveis de natureza emocional e não pela presença de simulação ou esforço insuficiente (Lee et al., 2000). Os diferentes estudos têm demonstrado que uma pontuação <45 no Ensaio 2 e de Reconhecimento do TOMM não pode ser atribuída à depressão, independentemente da sua severidade (Rees et al., 2001; Yanez, Fremouw, Tennant, Strunk, & Coker, 2006; O’Bryant, Finlay e Jile, 2007). O’Bryant, Finlay e Jile (2007) concluem ainda que o mesmo se verifica em relação à ansiedade.

No que diz respeito a adultos idosos, Lee e colaboradores (2000) desenvolveram um estudo para determinar em que medida os resultados no 15-IMT e no DCT (que requerem velocidade mental, capacidades visuo espaciais e perceptivas, memória visual, supostamente afectadas nesta população) seriam ou não afectados pelo quadro depressivo, bem como até que ponto é que a severidade da patologia teria impacto no desempenho. Os dados obtidos permitiram concluir que a depressão (independentemente do grau de severidade) não tem influência no desempenho nestes TVS. À mesma conclusão chegou o estudo de Ashendorf e colaboradores (2004) em relação ao TOMM. Numa amostra de 197 adultos idosos da comunidade, com idades compreendidas entre os 55 e os 75 anos, verificaram que o TOMM não é só insensível à presença de depressão (avaliado pelo *Beck Depression Inventory* - BDI) como também à presença de ansiedade, avaliado pelo *State-Trait Anxiety Inventory* (STAI).

Estudos prévios evidenciam que o estado de ansiedade e depressão comórbidos podem influenciar o desempenho em testes neuropsicológicos (Kizilbash, Vanderploeg & Curtiss, 2002), o que não se verifica nos TVS.

II - Objectivos

O objectivo geral desta investigação consiste na validação de um protocolo de avaliação de “simulação”, esforço reduzido/insuficiente ou exagero de sintomas em amostras de adultos idosos. Como objectivos mais específicos enumera-se: i) avaliar padrões de desempenho dos 3 grupos (comunidade, declínio cognitivo ligeiro, grupo instruído a “simular”); ii) analisar os valores de especificidade e sensibilidade de cada medida, de forma a assegurar validade e utilidade do protocolo; iii) perceber a influência

de variáveis demográficas e emocionais no desempenho em cada medida de simulação ou esforço reduzido/insuficiente; iv) analisar a possibilidade da utilização do Ensaio 1 do TOMM como medida de rastreio breve, a ser completada pelo recurso a outras medidas, caso haja evidência de simulação ou esforço insuficiente; v) verificar se existem diferenças entre o grupo de sujeitos instruídos a simular e os restantes grupos.

Tendo em conta estes objectivos e o que a literatura tem vindo a evidenciar, colocam-se as seguintes hipóteses de estudo:

a) Existem diferenças estatisticamente significativas entre grupos de idosos da comunidade e com diagnóstico de DCL em comparação com o grupo instruído a simular. Tendo em conta o pressuposto base da construção dos TVS relativamente a sua sensibilidade à simulação de défices e insensibilidade a variáveis sociodemográficas e défices reais (Tombaugh, 1997; Ashendorf et al., 2004) é esperado que os sujeitos com diagnóstico de DCL e da comunidade tenham desempenhos semelhantes e os sujeitos do grupo experimental tendam a apresentar resultados inferiores aos pontos de corte propostos nos diferentes índices alvo de estudo.

b) Variáveis sócio-demográficas como idade, escolaridade e género não interferem nos resultados obtidos nos TVS, em cada um dos grupos.

c) O estado emocional (depressão/ansiedade) não influencia os resultados nos diferentes TVS.

d) As diferentes medidas de avaliação de simulação/exagero de sintomas têm valores de especificidade adequados (elevados), equivalente a taxas de falsos positivos reduzidas.

e) O Ensaio 1 do TOMM pode constituir uma versão abreviada, válida e útil, sendo para isso necessária a análise de pontos de corte sugeridos na literatura.

f) Cada instrumento dá o seu contributo na identificação de simulação ou exagero de sintomas.

III - Metodologia

1. Amostra

As amostras do presente trabalho encontram-se caracterizadas no Quadro 1, sendo constituídas por um grupo de controlo (n=31), um grupo

clínico (n=17) e um grupo instruído a simular (n=29).

Mais especificamente, o grupo da comunidade tem, em média, 65,65 anos ($dp=10.045$) e 7,23 ($dp=3.603$) anos de escolaridade. O grupo com DCL tem em média 63,29 anos ($dp=7.753$) e 4,18 anos de escolaridade ($dp=1.286$). Já o grupo instruído a simular apresenta uma média de idades de 59,10 anos ($dp=3.678$) e 8,69 de escolaridade ($dp=3.983$).

Quadro 1: Caracterização da amostra para as variáveis sociodemográficas

n=77		Amostra da Comunidade (n=31)	Grupo Clínico (DCL) (n=17)	Amostra instruída a simular (n=29)
Género	Masculino	16 (51.6%)	2 (11.8%)	19 (65.5%)
	Feminino	15 (48.4%)	15 (88.2%)	10 (34.5%)
Idade		65.65 ± 10.045 (Min=55; Máx=86)	63.29 ± 7.753 (Min= 55; Máx = 81)	59.10 ± 3.678 (Min= 55; Máx = 70)
	55-64	n= 19	n= 11	n= 27
	65-74	n= 5	n= 4	n= 2
	+75 Anos	n= 7	n= 2	n= 0
Nível de escolaridade	0-4	13 (41.9%)	16 (94.1%)	9 (31.0%)
	5-9	9 (29.0 %)	1 (5.9%)	8 (27.6%)
	10-12	6 (19.4%)	0 (0%)	8 (27.6%)
	>12	3 (9.7%)	0 (0%)	4 (13.8%)
Estado Civil	Solteiro	0 (0%)	1 (5.9%)	1 (3.4%)
	Casado	24 (77.4%)	14 (82.4%)	25 (86.2%)
	Divorciado	1 (3.2%)	1 (5.9%)	2 (6.9%)
	Viúvo	6 (19.4%)	1 (5.9%)	1 (3.4%)
Situação Laboral	Reformado	20 (64.5%)	10 (58.8%)	15 (51.7%)
	Activo	11 (35.5%)	7 (41.2%)	14 (58.3%)

Os métodos de recrutamento e os respectivos critérios de inclusão e exclusão são descritos de seguida.

Amostra da comunidade. Este grupo foi constituído por sujeitos da comunidade residentes no concelho de Santa Maria da Feira (amostra de conveniência). Os critérios de inclusão foram: idade igual ou superior a 55 anos, ausência de evidência de declínio cognitivo (corroborada no ACE-R, de acordo com dados normativos por idade e escolaridade) e consentimento informado para participar voluntariamente no estudo. Os critérios de exclusão foram: evidência de abuso de substâncias; problemas psiquiátricos, neurológicos que possam induzir défices cognitivos; défices sensoriais não corrigidos (visuais/auditivos) e problemas de linguagem (compreensão) que inviabilizem aplicação do protocolo; presença de demência de acordo com

os critérios do DSM-IV-TR; e existência de motivos para simulação.

Amostra de sujeitos com Declínio Cognitivo Ligeiro. Foram seleccionados tanto da população da Comunidade como do Centro Hospitalar e Psiquiátrico de Coimbra – Unidade Sobral Cid (Consulta do Serviço de Psicogeriatría). Os critérios de inclusão incluíram: idade superior a 55 anos, presença de DCL de acordo com os critérios de diagnóstico propostos recentemente por Albert e colaboradores (2011)⁴. Os critérios de exclusão são os mesmos anteriormente mencionados.

Amostra instruída a simular: Este grupo foi recolhido na comunidade e é constituído por sujeitos com idade superior a 55 anos, a quem foram dadas instruções para simularem défices mnésicos e sintomatologia depressiva, que teriam repercussões na actividade profissional (cf. Anexo 3). Os critérios de inclusão e exclusão são sobreponíveis aos da amostra da comunidade.

2. Instrumentos

2.1 Questionário de dados demográficos. Conjunto de questões relativas a variáveis e características sociodemográficas.

2.2 Addenbrooke's Cognitive Examination (ACE-R; Mioshi, Dawson, Mitchell et al., 2006; Firmino, Simões, Pinho et al., 2008). Instrumento de rastreio cognitivo útil na detecção de demência ligeira e na diferenciação entre doença de Alzheimer e demência frontotemporal. Este teste inclui os seguintes domínios: Atenção e Orientação; Memória; Fluência; Linguagem; e Aptidões Visuo-espaciais.

2.3 Inventário de Avaliação Funcional de Adultos Idosos (IAFAI; Sousa, Simões, Pires, Vilar, & Freitas, 2008). O IAFAI é um instrumento de avaliação funcional, constituído pelos seguintes módulos: Avaliação da Capacidade Funcional Global; Actividades Básicas de Vida Diária; Actividades Instrumentais Familiares de Vida Diária; Actividades Instrumentais Avançadas de Vida Diária. Pode ser respondido pelo doente

⁴ Os critérios de diagnóstico adoptados consistem em: alteração cognitiva reportada pelo paciente, informador, ou clínico; evidência objectiva de défice em pelo menos um domínio cognitivo; dificuldades nas actividades mais complexas de vida diária; ausência de demência.

ou pelo cuidador. O respondente deve reportar-se ao último mês e considerar um gradiente de nível de dificuldade.

2.4 Test of Memory Malinger (TOMM; Tombaugh, 1996; Mota et al, 2009; Fernandes, 2009). O TOMM é uma medida de escolha forçada onde são demonstrados aos sujeitos 50 imagens durante 3 segundos cada, com um intervalo de 1 segundo. Imediatamente depois são apresentados pares de imagens com a imagem prévia e uma nova tendo o sujeito que dizer qual a imagem previamente apresentada. É dado *feedback* após cada resposta. O teste é composto por dois Ensaio de Aprendizagem e por um Ensaio de Retenção (após 10 minutos). É atribuído 1 ponto por cada resposta correcta, podendo a pontuação total variar entre 0 e 50.

2.5 Dot Counting Test (DCT; Rey, 1941; Boone et al., 2002). O teste consiste em dois grupos de cartões. Os primeiros 6 cartões contêm 11, 19, 15, 23, 27 e 7 pontos, distribuídos de forma aleatória. Os restantes 6 cartões contêm 12, 20, 16, 24, 28 e 8 pontos agrupados. Os cartões são apresentados por esta ordem e os sujeitos têm de contar os pontos o mais depressa possível, podendo usar o dedo ou lápis para contar se desejarem. São anotadas as respostas e o tempo dispendido para cada cartão. Os pontos agrupados são mais fáceis de contar pelo que se espera que, quando os sujeitos cooperam, irão contar os pontos agrupados mais rapidamente que os não agrupados. Os resultados obtidos incluem o total de erros, média de tempo para contar pontos agrupados, média de tempo usada para contar pontos não agrupados, razão entre estes dois últimos (média de tempo para contar pontos não agrupados a dividir pela média do tempo para contar os pontos agrupados), e resultado total.

2.6 b-Test (Boone et al., 2000). A tarefa do sujeito é rodear todas as letras “b”, num conjunto de estímulos constituídos pelas letras “b”, “d”, “p” e “q”. À medida que se vão apresentando os estímulos as letras vão-se tornando mais pequenas para que a tarefa pareça mais difícil e induza os sujeitos a simular défices. É dado aos participantes a instrução “Circule todas as letras “b” nestas páginas o mais rápido que conseguir. As letras que terá de circular devem ser exactamente iguais a esta (mostrar exemplo)”. As

medidas possíveis são: o número de erros de comissão e de omissão, tempo total para completar a tarefa e resultado total (Boone et al., 2000).

2.7 Rey 15-Item Memory Test (Rey, 1964; Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002; Simões, Sousa, Duarte, Firmino, Pinho, Gaspar, et al., 2010). Este é um teste simples, constituído por 15 itens (3 colunas e 5 linhas, com 3 caracteres em cada linha) e inclui, para além do Ensaio de Evocação Imediata, um Ensaio de Reconhecimento (apresentação dos 15 itens do cartão estímulo original e 15 distractivos, semelhantes aos itens do cartão inicial). Na realidade, o teste só requer a memorização de 5 elementos, uma vez que os itens podem ser agrupados em categorias. Para além da consideração do número total de itens que foram correctamente reproduzidos no Ensaio de Evocação Imediata, os indicadores relativos ao Ensaio de Reconhecimento contemplam: (1) o número de itens correctamente identificados/reconhecidos; (2) o número de falsos positivos (itens incorrectamente assinalados); e (3) a pontuação combinada.

2.8 Structured Inventory of Malingering Symptomatology (SIMS; Smith & Burger, 1997; Widows & Smith, 2005; Simões et al., 2010). O SIMS é um inventário de auto-resposta, composto por 75 itens, utilizado para detectar a presença de comportamentos de simulação em contextos clínicos e forenses. Este instrumento é constituído por 5 escalas (15 itens cada), para detecção de sintomas psicopatológicos, afectivos e também cognitivos. As escalas são: *Psicose*, *Défice Neurológico*, *Perturbações Mnésicas*, *Capacidade Intelectual Reduzida*, *Perturbações Afectivas*. Cada item é pontuado com 0 ou 1 pontos, sendo que a pontuações total varia de 0 a 75. Uma pontuação total acima do ponto de corte de 16 é indicativa de possível simulação ou esforço insuficiente.

2.9 Subteste Vocabulário (WAIS-III; Wechsler, 1998, 2008). Subteste da WAIS-III. Avalia a capacidade intelectual verbal e contribui para o cálculo do QI verbal da escala. A tarefa do sujeito consiste em definir oralmente um conjunto de palavras. É considerado uma medida da inteligência verbal, cristalizada e pré-mórbida uma vez que não declina com o aumento da idade. As pontuações por item variam entre 0 e 2 pontos, e o resultado total entre 0 e 66 pontos.

2.10 Subteste Memória de Dígitos (WAIS-III; Wechsler, 1997, 2008). Subteste da WAIS-III. Requer que os sujeitos repitam várias sequências de números de diferente extensão, em sentido directo (envolvidos processos atencionais) e, posteriormente, em sentido inverso (envolvida a memória de trabalho, pois requer o armazenamento durante um período curto de tempo para depois serem trabalhados e evocados em sentido inverso). O Índice de Fiabilidade da Memória de Dígitos e o Índice Memória de Dígitos – resultados padronizados de acordo com a idade (resultado total) são medidas referenciadas para detecção de simulação e esforço insuficiente. São ainda alvo de análise os tempos de resposta para os 3 e 4 dígitos em sentido directo.

2.11 Subteste Memória Lógica I e II (WMS-III, Wechsler, 1997, 2008). Avalia a memória episódica e semântica através da apresentação verbal de 2 histórias curtas ao sujeito. Estas devem ser evocadas logo de seguida (a segunda história é lida e evocada 2 vezes). Após um intervalo de 25 a 30 minutos procede-se ao ensaio de evocação diferida das duas histórias, ensaios que precedem a tarefa de reconhecimento. Neste subteste o Índice de Respostas Infrequentes é usado como medida de detecção de simulação ou esforço insuficiente.

2.12 Escala de Depressão Geriátrica – 30 itens (GDS-30; Yesavage, Brink, Rose et al., 1983; Barreto, Leuschner, Santos & Sobral, 2008; Simões, Firmino, Sousa et. al., 2010). Esta escala, composta por 30 itens de resposta dicotómica (Sim/Não), avalia presença de sintomas afectivos e comportamentais da depressão.

2.13 Geriatric Anxiety Inventory (GAI; Pachana et al., 2006; versão portuguesa: Ribeiro, Paúl, Simões, & Firmino, 2011). Escala constituída por 20 itens. Avalia a intensidade de sintomas ansiosos, num formato de resposta do tipo “concordo/discordo”.

3. Procedimentos

Antes da recolha dos dados procedeu-se ao pedido de autorização das Comissões de Ética envolvidas, nomeadamente do Centro Hospitalar e Universitário de Coimbra.

Previamente à aplicação do protocolo os participantes foram informados acerca dos objectivos da avaliação, confidencialidade dos dados, bem como da possibilidade de desistência da investigação, assinando o consentimento informado. A aplicação do protocolo demorou sensivelmente entre 90 a 120 minutos. Ao grupo instruído a simular as instruções foram fornecidas após aplicação do ACE-R, GDS-30, GAI e IAFAI.

Foram alvo de análise os seguintes índices: pontuação no Ensaio 2 e no Ensaio de Reconhecimento do TOMM; total de erros, tempo para contar pontos agrupados e não agrupados, média de tempo para contar pontos agrupados e não agrupados, razão da média de tempo usada para contar pontos não agrupados por agrupados e resultado combinado (*E-score*) do *Dot Counting Test*; tempo total, erros de omissão e comissão no *b-test*; resultado da evocação imediata e resultado combinado do 15-IMT; resultado total do SIMS; Resultado padronizado, FMD, 3 e 4 dígitos cronometrados em sentido directo do subteste Memória de Dígitos; Índice de Respostas Infrequentes (*Rarely Missed Index - RMI*) do Subteste Memória Lógica; e o índice Vocabulário menos Memória de Dígitos.

Todas as análises estatísticas foram realizadas com o auxílio do programa *Statistical Package for the Social Sciences* (SPSS – versão 17.0). Um valor de $p < .05$ foi considerado estatisticamente significativo. Relativamente à comparação dos grupos, dado o reduzido tamanho das amostras em estudo e o facto dos resultados nos testes não estarem distribuídos de forma normal, foram utilizadas comparações não paramétricas, nomeadamente o U de Mann-Whitney. Para perceber a influência das variáveis sociodemográficas e depressão/ansiedade no desempenho nos diferentes grupos utilizaram-se correlações de Pearson para variáveis contínuas (idade e escolaridade) e comparações Mann-Whitney para a variável categorial género.

IV - Resultados

1. Comparação do desempenho dos grupos nas diferentes medidas

Procedeu-se à análise comparativa do desempenho nos diferentes grupos em estudo procurando-se analisar se os sujeitos com DCL tinham

desempenhos semelhantes aos da comunidade e se estes diferiam do grupo instruído a simular.

Contrariamente ao esperado, verificou-se que a amostra da comunidade e o grupo clínico têm desempenhos semelhantes apenas em algumas medidas (cf. Quadro 2), nomeadamente nos dois ensaios do TOMM, no índice 4 dígitos cronometrados da Memória de Dígitos, total de erros do *Dot Counting Test* e no índice Vocabulário menos Memória de Dígitos. Os restantes índices têm valores de $p < .05$, correspondente a diferenças estatisticamente significativas entre os grupos em comparação.

Quadro 2: Análise das diferenças entre grupos nas diferentes medidas de simulação/esforço insuficiente

Índices	Comunidade <i>versus</i> DCL	Comunidade <i>versus</i> Instrução
RMI	U=124.500, p=0.002	U=243.000, p=.002
TOMM: Ensaio 2	U=246.500, p=.661	U=173.500, p=.000
TOMM: Ensaio de Reconhecimento	U=240.500, p=.535	U=150.500, p=.000
Memória Dígitos (MD): Pontuação Total	U=103.000, p=.000	U=330.500, p=.076
MD: FMD	U=119.000, p=.002	U=365.000, p=.206
MD: Tempo 3 dígitos	U=147.500, p=.012	U=396.500, p=.433
MD: Tempo 4 dígitos	U=178.000, p=.258	U=399.000, p=.585
SIMS Total	U=111.500, p=.001	U=9.000, p=.000
b-test: Tempo total resposta	U=152.500, p=.017	U=404.500, p=.515
<i>b-test</i> : Total erros comissão	U=139.000, p=.005	U=182.500, p=.000
<i>b-test</i> : Total erros omissão	U=113.500, p=.001	U=258.500, p=.005
<i>b-test</i> : Resultado Total (E-score)	U=135.000, p=.006	U=195.00, p=.000
<i>Dot Counting Test</i> (DCT): Total Erros	U=211.500, p=.216	U=449.000, p=.993
DCT: Tempo total pontos agrupados	U=89.500, p=.000	U=410.000, p=.559
DCT: Tempo total pontos não agrupados	U=157.000, p=.022	U=445.000, p=.947
DCT: Média para contar pontos agrupados	U=108.000, p=.001	U=410.000, p=.559
DCT: Média para contar pontos não agrupados	U=151.000, p=.015	U=445.000, p=.947
DCT: Razão	U=145.000, p=.011	U=369.000, p=.371
DCT: Resultado total (E-Score)	U=143.000, p=.009	U=446.500, p=.964
Rey 15-IMT: Evocação Imediata	U=135.000, p=.005	U=395.000, p=.408
Rey 15-IMT: Resultado Combinado	U=109.500, p=.001	U=448.500, p=.998
Vocabulário – MD	U=204.000, p=.193	U=449.000, p=.994

Relativamente às comparações entre o grupo da comunidade e o instruído a simular verificaram-se diferenças em todos os índices, com excepção dos do DCT, *Rey 15-Item Memory Test*, Memória de Dígitos, Vocabulário menos Memória de Dígitos e tempo total de resposta do *b-test*.

A análise comparativa dos desempenhos dos sujeitos com DCL amnésico e não amnésico permitiu concluir que não existem diferenças significativas nos diferentes índices ($p > 0.05$) (cf. Anexo 4).

2. *Efeito das variáveis idade, escolaridade e género no desempenho nos testes de simulação e/ou esforço insuficiente (para cada um dos grupos)*

O efeito da idade, escolaridade e género foi examinado em cada um dos grupos em estudo para os diferentes índices passíveis de serem obtidos no presente protocolo (cf. Quadro 3 e 4). Contrariamente ao esperado relativamente à insensibilidade dos diferentes testes às variáveis sociodemográficas, podemos observar que os únicos índices que não sofrem influência⁵ da idade, escolaridade e género nos 3 grupos são o RMI, 3 e 4 dígitos cronometrados da MD, SIMS total, Erros de omissão do *b-test* e tempo total para contar pontos não agrupados do DCT.

No grupo de controlo (cf. Quadro 3), verificou-se influência (baixa) da idade e escolaridade no Ensaio 2 do TOMM, sendo que no Ensaio de Reconhecimento esta influência se verifica apenas em relação à escolaridade. Pode-se também verificar que quanto maior a idade menor a pontuação no índice FMD e maior o tempo de resposta no *b-test* (associações baixas). Ainda em relação a este último teste, verificou-se que a escolaridade tem um efeito moderado negativo na pontuação total. Relativamente ao DCT verificou-se influência da idade e escolaridade (moderada e positiva) no número total de erros. Observaram-se diferenças de género no tempo total para contar pontos agrupados e na média para contar pontos agrupados (mulheres demoram mais tempo). Pode-se ainda verificar neste instrumento um efeito moderado positivo da idade na pontuação total. No 15-IMT observou-se uma influência significativa moderada da idade e escolaridade na Evocação Imediata e Resultado Combinado.

Da análise dos desempenhos através de estatísticas descritivas, foi possível verificar que, apesar de se identificar esta influência da idade e

⁵ Os valores de correlação variam entre -1 e 1. Se o sinal é negativo significa que as variáveis se relacionam em sentido contrário e se positivo que estas relacionam-se no mesmo sentido. Um r entre 0-0.2 é considerada uma associação muito baixa, entre 0.2-0.4 baixa, 0.4-0.6 moderada, 0.6-0.8 alta e maior que 0.8 muito alta. Em relação à significância, quando $p < .05$ considera-se uma associação significativa.

escolaridade, os sujeitos continuam a ter desempenhos esperados no TOMM (nenhum sujeito obtém pontuações inferiores ao ponto de corte no Ensaio 2 no Ensaio de Reconhecimento), FMD (com exceção de um sujeito que teve 6 pontos), resultado total do *b-test* (excepto 4 sujeitos), número de erros e tempo para contar pontos não agrupados do DCT, e resultado total (apenas um sujeito teve 17 pontos).

Quadro 3: Efeito das variáveis sociodemográficas no desempenho no grupo de controlo e DCL

Índices	Controlo (n=31)			DCL (n=17)		
	Idade	Esc.	Gen.	Idade	Esc.	Gen.
RMI	r= -.066	r= -.029	U=73.5000	r= -.257	r=.003	U=12.000
TOMM						
Ens. 2	r= -.374*	r= .371*	U=116.000	r= .149	r= -.129	U=9.000
Ens. Rec.	r= -.350	r= .375*	U=113.500	r= -.402	r= .196	U=9.000
MD						
Total	r= .028	r= .119	U=102.000	r= -.256	r= .654**	U=14.500
FMD	r= -.368*	r= .155	U= 91.500	r= -.469	r= .671**	U=7.500
3 díg.	r= .043	r= -.22	U=111.000	r= -.132	r= -.212	U=8.000
4 díg.	r= .147	r= -.148	U= 79.500	r= .464	r= -.165	U=4.000
SIMS Total	r= .163	r= -.159	U=80.500	r= -.209	r= -.104	U=4.000
b-test						
Tempo	r= .380*	r= -.50	U=91.500	r= .670**	r= -.515*	U=4.000
E. Com.	r= .098	r= -.344	U= 97.000	r= .346	r= -.281	U=5.000
E. Omissão	r= .164	r= -.338	U=107.500	r= .379	r= -.274	U=3.000
Total	r= .219	r=-.435*	U=95.500	r= .360	r= -.308	U=4.000
Dot Counting Test						
Erros	r= .358*	r=-.446*	U=101.500	r= .376	r= -.088	U=11.500
Tempo agrup.	r= .174	r= -.045	U=66.000*	r= .455	r= -.340	U=8.000
Tempo não ag.	r= .518**	r= .138	U= 81.000	r= .326	r= -.394	U=6.000
Média agrup.	r= .170	r= -.035	U=66.000*	r= .487 *	r= -.312	U=.000*
Razão	r= .269	r= -.058	U= 85.000	r= -.330	r= .021	U=4.000
Total	r= .403*	r= .062	U= 71.000	r= .531 *	r= -.443	U=3.000
Rey 15-Item Memory Test						
Evoc. lmed.	r= -.436*	r= .568**	U= 115.000	r= -.202	r=.380	U=9.500
Res. Comb.	r=-.512**	r= .613**	U= 119.000	r= -.472	r= .336	U=8.000
Voc – MD	r= .037	r= .270	U= 110.000	r= .429	r= -.466	U=14.000

*p<.05; **p<.01

No grupo clínico, os resultados das correlações permitem verificar um efeito elevado positivo da escolaridade na pontuação total e FMD da MD⁶.

⁶ Sendo estas provas de memória, a educação tem uma influência nos resultados que não é surpreendente. Vários estudos têm demonstrado um declínio cognitivo mais rápido em indivíduos menos escolarizados.

Relativamente ao *b-test* observou-se influência da idade (alta) e escolaridade (moderado) no tempo de resposta. No DCT verificou-se que as mulheres demoram mais tempo em média para contar os pontos agrupados bem como efeito (moderado) da idade neste índice e na pontuação total.

Em relação ao grupo instruído a simular (cf. Quadro 4), não se verificou influência de idade em nenhum dos parâmetros avaliados.

Quadro 4: Efeito das variáveis sociodemográficas no desempenho no grupo instruído a simular

Índices	Instrução (n=30)		
	Idade	Esc.	Género
RMI	r= .226	r= -.257	U=87.000
TOMM			
Ens. 2	r=-.073	r=-.327	U=51.500*
Ens. Rec.	r=-.032	r=-.246	U=66.500
MD			
Total	r=-.0131	r= -.120	U=61.500
FMD	r= -.205	r= -.077	U=61.000
3 díg.	r= -.140	r= .225	U=65.000
4 díg.	r= .059	r= -.025	U=69.000
SIMS Total	r= -.092	r= .320	U=81.000
b-test			
Tempo	r= .228	r= .130	U=63.500
E. Com.	r= -.164	r= .401*	U=66.500
E. Omissão	r= -.133	r= .341	U=76.000
Total	r= -.203	r= .457*	U=60.500
Dot Counting Test			
Erros	r= -.346	r=-.156	U=91.000
Tempo agrup.	r= .008	r= -.104	U=35.000**
Tempo não ag.	r= .043	r= .156	U=61.000
Média agrup.	r= .009	r= -.104	U=34.000**
Razão	r= -.065	r= .419*	U=60.500
Total	r= -.180	r= .117	U=46.000*
Rey 15 item Memory Test			
Evoc. Imed.	r= -.198	r= .205	U=70.000
Res. Comb.	r= -.137	r= .194	U=73.500
Voc – MD	r= -.051	r= .443*	U=62.000

*p<.05; **p<.01

A escolaridade apenas tem efeito significativo e moderado nos erros

Deste modo, sujeitos com uma Reserva Cognitiva mais baixa manifestam maior risco de desenvolvimento de demência, tendo a escolaridade e nível ocupacional um efeito protector no declínio cognitivo (Pires, Simões & Firmino, 2010).

de comissão e no total do *b-test*, na razão do DCT, bem como no índice Vocabulário menos Memória de Dígitos. Por outro lado, verificaram-se diferenças de género no desempenho no ensaio 2 do TOMM, no tempo para contar pontos agrupados, média de tempo dos pontos agrupados e total do DCT (aparentemente são as mulheres que “simulam” mais).

3. *Análise da influência das alterações de humor nos desempenhos nos diferentes resultados dos TVS*

Para avaliar a relação entre as diferentes medidas e a sintomatologia afectiva foram efectuadas análises de correlação, para cada um dos grupos. Relativamente à influência da sintomatologia depressiva (medida pela GDS-30) observaram-se correlações estatisticamente significativas ($p < 0.05$) no resultado total do SIMS no grupo de controlo ($r = 0.449$; $p = 0.011$) e DCL ($r = 0.726$; $p = 0.001$). Esta associação é moderada e alta, respectivamente. Já no grupo instruído a simular, estas diferenças encontram-se no tempo total para contar pontos agrupados ($r = 0.390$; $p = 0.037$), na média de tempo para contar pontos agrupados ($r = 0.391$; $p = 0.036$) e no valor da razão ($r = -.530$; $p = .003$), índices do DCT. Esta última associação é moderada, sendo as restantes associações fracas.

Em relação à influência da ansiedade (medida através do GAI), verificou-se, de novo, este efeito significativo e moderado no grupo de controlo ($r = .624$; $p < 0.001$) e DCL ($r = .587$; $p = 0.013$) no resultado total do SIMS. No grupo instruído a simular observaram-se diferenças no tempo total para contar pontos agrupados ($r = 0.435$; $p = 0.018$), na média de tempo para contar pontos agrupados ($r = 0.436$; $p = 0.018$) e no valor da razão ($r = -.524$; $p = .003$), também com associações moderadas.

Relativamente à influência da depressão/ansiedade no grupo de controlo e DCL no SIMS verificou-se que os sujeitos têm pontuações significativamente mais elevadas nas subescalas *perturbações afectivas*, *défice neurológico* e *perturbações mnésicas* (cf. Anexo 5).

4. *Análise dos valores de especificidade (grupo da comunidade e DCL) e sensibilidade (instrução) dos diferentes índices*

Verificou-se que, apesar dos valores de sensibilidade baixos na amostra instruída a simular (cf. Quadro 5), todos os sujeitos são identificados em pelo menos uma destas medidas, o que corrobora a necessidade de recurso a diferentes instrumentos para se concluir a possível presença de “simulação” (cf. Tombaugh, 1997; Boone et al., 2000; Rosenfeld et al., 2010; Simões et al., 2010; Rudman et al., 2011).

Quadro 5: Valores de especificidade e sensibilidade dos diferentes TVS

	Ponto Corte	Comunidade	DCL	Instrução
RMI	≤136	90.32%*	94.12%*	48.28%
TOMM				
Ensaio 2	<45	100%*	100%*	62.07%
Ensaio de Reconhecimento	<45	100%*	94.12%*	62.07%
Memória de Dígitos				
Pontuação Total	≤5	100%*	94.12%*	10.34%
FMD	≤6	96.77%*	70.59%	13.79%
Tempo 3 dígitos	>2s	93.55%*	70.59%	17.24%
Tempo 4 dígitos	>4s	83.87%*	58.82%	24.14%
SIMS Total	>16	93.55%*	58.82%	96.55%*
b-test				
Tempo total	>12	80.65%*	47.06%	37.93%
Total erros comissão	>2	74.19%	41.18%	65.52%
Total erros omissão	>40	93.55%*	70.59%	27.59%
Resultado Total	≥160	87.1%*	64.71%	48.28%
Dot Counting Test				
Total Erros	>3	100%*	100%*	6.90%
Tempo total pontos agrup.	>130s	100%*	100%*	0%
Tempo t. pontos não agrup.	>180s	100%*	100%*	0%
Média contar pontos agrup	>7s	100%*	84.24%*	0%
Tempo contar pontos agrupados>não agrupados		100%*	88.24%*	0%
Razão	≤1.5	87.1%*	47.06%	17.24%
Resultado total	≥17	96.77%*	100%*	6.90%
Rey 15-Item Memory Test				
Evocação Imediata	<9	83.87%*	64.71%	3.45%
Evocação Imediata	<6	93.55%*	88.24%*	0%
Resultado Combinado	<20	74.19%	35.29%	13.79%
Voc – MD	>5	96.77%*	100%*	10.34%

*Valores > 75% são considerados adequados

Apenas o SIMS possui valores de sensibilidade adequados, apresentando-se como a única medida com elevada capacidade de detectar

os “simuladores” quando de facto o são, sendo acompanhado de baixas taxas de falsos positivos, mas apenas na amostra da comunidade. Em sujeitos com algum declínio cognitivo, a especificidade desta medida é inferior, o que poderá indicar que valores inferiores ao ponto de corte poderão levar à inferência de que os sujeitos estão a simular quando não o estão.

O índice RMI, o TOMM, Pontuação total da MD, DCT (excepto valor da razão), Evocação Imediata (<6) do 15-IMT e Vocabulário menos Memória de Dígitos, são os índices que continuam com taxas de falsos positivos reduzidas na amostra da comunidade e com DCL e como tal aparentemente insensíveis ao declínio cognitivo (e válidos para estes grupos). Ressalva-se que o ajuste do ponto de corte do Ensaio da Evocação Imediata do 15-IMT compromete a capacidade deste ensaio detectar os “simuladores” (cf. Dean et al., 2009).

5. Análise crítica dos pontos de corte propostos para o Ensaio 1 do TOMM.

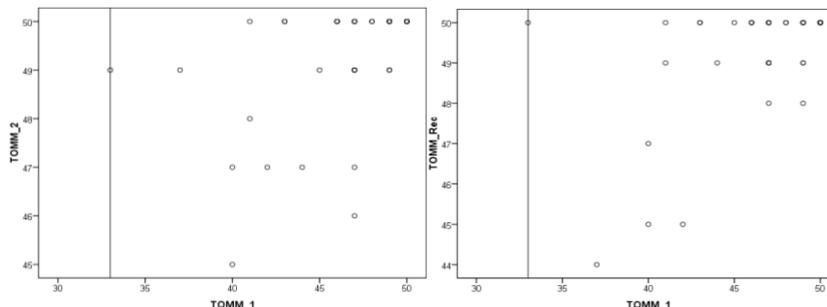
O objectivo deste trabalho é também verificar se os pontos de corte apresentados na literatura (cf. Greve & Bianchini, 2006; O’Byrant et al., 2007; Bauer et al., 2007; O’Byrant et al., 2008; Fernandes, 2009; Denning, 2012) são adequados para detectar “simulação” ou exagero de sintomas. Procurou-se ainda analisar se, tal como referido por O’Byrant e colaboradores (2008), os sujeitos com pontuações ≥ 45 no Ensaio 1 continuam com o mesmo desempenho nos restantes ensaios, o que viabiliza a não administração dos ensaios seguintes.

De acordo com o que é possível observar no gráfico 1, os dados corroboram o ponto de corte ≥ 33 definido por Fernandes (2009) como capaz de classificar correctamente 100% dos não “simuladores” (correspondente à pontuação mínima da amostra da comunidade e DCL). Os pontos de corte propostos por Hilsabeck e colaboradores (2011) de ≤ 41 ou por Denning (2012) de ≤ 40 são refutados, verificando-se que nesta base alguns sujeitos seriam classificados erradamente como “simuladores”.

Verifica-se ainda que os sujeitos que têm pontuações superiores a 33 no Ensaio 1 não têm resultados inferiores a 45 nem no Ensaio 2 nem no

Ensaio de Reconhecimento o que corrobora que este ponto de corte poderá ser indicativo de esforço suficiente.

Gráfico 1: Gráficos de dispersão (amostra comunidade e DCL) dos valores no Ensaio 2 e Ensaio de Reconhecimento em função do Ensaio 1 do TOMM

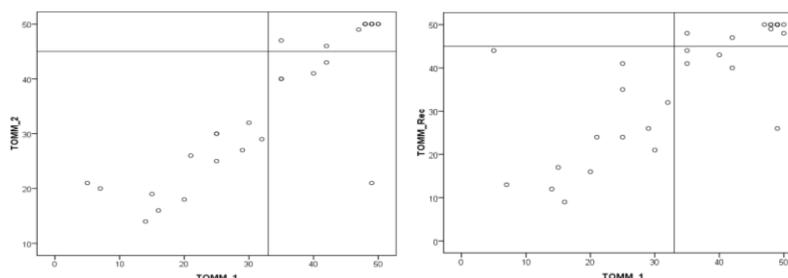


É ainda possível constatar que os sujeitos que apresentam resultados iguais ou superiores a 45 no Ensaio 1 continuam com estes desempenhos nos restantes ensaios, corroborando estudos prévios (cf. O’Byrant et al., 2008; Fernandes, 2009) que sugerem que estas pontuações poderão indicar não ser necessária a aplicação dos restantes ensaios.

Contudo, surge a questão de saber se uma pontuação inferior a este ponto de corte também detecta correctamente os sujeitos “simuladores”. Através da análise do gráfico 2, pode-se observar que todos os sujeitos que tiveram pontuações inferiores a 33 (Fernandes, 2009) ou 25 (Hilsabeck et al., 2011) no Ensaio 1 continuam a ter desempenhos < 45 no Ensaio 2 e no Ensaio de Reconhecimento. Tais resultados são indicativos de que pontuações no TOMM <33 ou mesmo 25 sugerem a possibilidade de simulação/esforço insuficiente. Convém contudo realçar que, não é apenas pelo facto dos sujeitos terem pontuações inferiores neste ensaio que são ou não “simuladores”. Esta decisão não se pode restringir apenas a este ensaio, até porque existem indivíduos da amostra instruída a simular que têm pontuações superiores a 33 e que continuam com desempenhos inferiores a 45 nos restantes ensaios. Além do mais, como já analisado previamente, o TOMM tem associada uma taxa de falsos negativos de 37.93%. De facto, no grupo de sujeitos instruídos a simular, alguns indivíduos têm desempenhos superiores a 33 (Fernandes, 2009), a 41 (Hilsabeck et al., 2011) ou a 40 (Denning, 2012) no Ensaio 1 e nos restantes ensaios têm pontuações sugestivas de exagero de sintomas. Verificou-se ainda que, contrariamente às conclusões de O’Byrant e colaboradores (2008), uma pontuação >45 no Ensaio 1 não significa que os sujeitos continuem com esses desempenhos

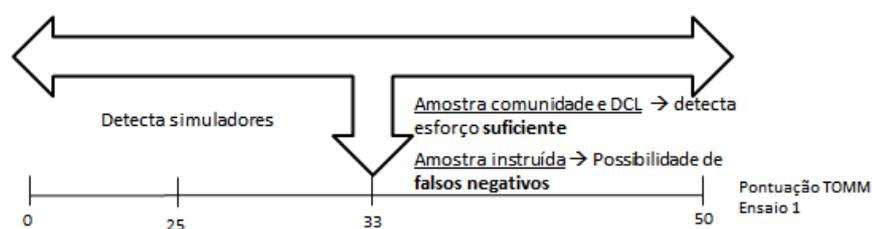
nos restantes ensaios. Esse padrão de respostas apenas se verificou na amostra da comunidade e DCL; no grupo instruído, é possível observar que existem sujeitos com desempenhos superiores a 45 no Ensaio 1 e desempenhos inferiores a este ponto de corte nos ensaios seguintes, bem como indivíduos com desempenhos superiores a 45 em todos os ensaios.

Gráfico 2: Gráficos de dispersão (amostra instruída a simular) dos valores no Ensaio 2 e no Ensaio de Reconhecimento em função do Ensaio 1 do TOMM



Em suma, através da representação (cf. Figura 1) poderemos assim concluir que, resultados inferiores a 33 pontos (resultado mínimo da amostra com DCL) no Ensaio 1 são indicativos de que os sujeitos estarão a “simular” ou a exagerar sintomas; contudo, pontuações superiores a este ponto de corte não permitem concluir que os sujeitos não estarão a “simular”. Deste modo, o ponto de corte proposto por Fernandes (2009) revela-se útil na detecção de simuladores, não menosprezando a possibilidade destes sujeitos obterem pontuações acima destes valores (falsos negativos).

Figura 1: Representação dos desempenhos dos sujeitos no TOMM



6. Contributo de cada medida na identificação de simulação ou esforço insuficiente

Os valores de especificidade e sensibilidade distintos de cada medida de simulação/esforço insuficiente colocam o problema de se poder aplicar um teste e este não detectar os “simuladores” quando estes o são ou, pelo contrário, sugerir que os sujeitos estão a “simular” e isso não ser verdadeiro.

Como foi possível observar nos resultados obtidos, os sujeitos quando simulam, não o fazem todos nos mesmos instrumentos (Rosenfeld et al., 2010). De facto, foi possível observar que cada indivíduo tem desempenhos inferiores aos pontos de corte em pelo menos uma medida de simulação ou esforço reduzido. Os únicos indicadores em que nenhum dos sujeitos simulou défices dizem respeito aos índices cronometrados do DCT. Neste instrumento, relativamente ao número total de erros, 2 sujeitos foram identificados por este índice; o valor da razão sinaliza 5 sujeitos; o resultado total do DCT identifica 2 sujeitos. Contudo, verifica-se que estes sujeitos detectados pelo DCT, são também assinalados noutras medidas sendo detectados em mais 1 a 12 medidas de simulação ou esforço insuficiente. O único sujeito que errou em apenas mais um índice além do valor da razão do DCT foi no resultado total do SIMS, associado a elevados valores de sensibilidade e valores relativamente adequados de especificidade.

Os valores nulos de sensibilidade do DCT parecem dever-se ao facto de os pontos de corte apresentados serem de tal forma exigentes que nenhum sujeito é captado por estes índices. Como tal, parece existir evidência que, face ao número reduzido de índices do DCT que possuem algum valor de sensibilidade, a sua aplicação não acrescenta evidência complementar para a distinção entre “simuladores” ou “honestos”. Deste modo, o recurso ao DCT torna-se dispensável, não fornecendo aparentemente validade adicional às restantes provas que compõem o presente protocolo. Procedeu ao ajuste dos pontos de corte deste instrumento e para tal verificou-se as pontuações máximas obtidas nestes índices pela amostra da comunidade e DCL (cf. Quadro 6).

Quadro 6: Amplitude de valores para os índices do DCT na amostra da comunidade, grupos com DCL e instruído a simular

	Comunidade	DCL	Instrução
Total Erros	0-3	0-2	0-8
Tempo total pontos agrup.	4.45-36.86	11.89-46.41	5.85-40.31
Tempo total pontos não agrupados	18.08-60.58	10.64-51.62	19.23-59.36
Média contar pontos agrupados	0.74-6.14	1.77-7.74	0.98-6.72
Razão	1.00-5.50	0.82-3.57	1.12-5.21
Resultado total	5-17	6-16	5-20

Com um ponto de corte >60 segundos para contar pontos agrupados, >70 para contar pontos não agrupados e >8 segundos no valor da média para

contar pontos agrupados, os valores de especificidade e sensibilidade continuam a ser de 100% e 0%, respectivamente. Com pontos de corte menos exigentes, estas medidas continuam a não ser capazes de detectar os “simuladores”. Com pontos de corte inferiores a estes, o valor de sensibilidade aumentaria mas comprometeria as taxas de falsos positivos.

V - Discussão

O exame sistemático da validade dos desempenhos nas avaliações neuropsicológicas é um aspecto de extrema importância e relevância mas nem sempre considerado pelos profissionais. Apesar da utilização dos TVS, o seu uso ainda não constitui uma rotina na prática clínica. O objectivo deste trabalho consistiu na validação de um protocolo de TVS em adultos idosos.

De acordo com a literatura e com os pressupostos base da construção dos TVS, as diferentes medidas seriam insensíveis a variáveis demográficas e afectivas, mas sensíveis à simulação de défices. Era assim esperado que os sujeitos da comunidade e com DCL tivessem desempenhos semelhantes e que estes por seu turno diferissem do grupo instruído a simular.

Os resultados do presente estudo são indicativos de que o grupo da comunidade e com DCL apenas têm desempenhos semelhantes, e congruente com outros estudos que referem a insensibilidade destas medidas ao declínio cognitivo, em alguns índices, nomeadamente no TOMM (Fernandes, 2009), 4 dígitos cronometrados, Vocabulário menos Memória de Dígitos (Dean et al., 2009) e total de erros do DCT (Boone et al., 2002a). Nos restantes, e contrariamente ao pressuposto base, a população clínica tem, em média, pior desempenho que o grupo de controlo (cf. Anexo 6). Os resultados obtidos no 15- IMT estão de acordo com os dados de Simões e colaboradores (2010) que verificaram uma necessidade de ajuste dos pontos de corte para a população mais idosa e com défice cognitivo. Já nos restantes índices das subescalas de Wechsler, nomeadamente no RMI, FMD e 3 dígitos cronometrados da MD, tais resultados também são justificáveis, visto serem medidas em que se observam piores desempenhos em sujeitos com alguns défices. Ainda contrariamente a resultados prévios verifica-se uma sensibilidade do DCT e *b-test* ao declínio cognitivo (Boone et al., 2000; Boone et al., 2002a), o que realça a necessidade de prudência quando se

aplicam estes instrumentos a estas populações, sob pena de os considerar erradamente como “simuladores”. Por outro lado, foi possível verificar diferenças significativas entre o grupo da comunidade e o instruído a simular em todas as medidas, com excepção dos índices da Memória de Dígitos, do DCT, 15-IMT, tempo total do *b-test* e do Vocabulário menos Memória de Dígitos. Relativamente às provas de Wechsler, mais especificamente na MD e Memória Lógica, verificou-se que os sujeitos praticamente não são identificados nestas medidas. Da percepção e *feedback* que ia sendo dado pelos participantes ao longo da aplicação dos protocolos, estes “simulavam” com mais facilidade nos TVS. Este padrão de respostas é uma consequência possível destes testes serem cognitivamente menos exigentes. Nestas medidas de memória os indivíduos procuravam também testar as suas capacidades. Nos TVS, já obtiveram maioritariamente desempenhos inferiores ao grupo da comunidade. Estes resultados corroboram o seguinte. Por um lado, nem todos os índices são à partida insensíveis ao declínio cognitivo (cf. Dean et al., 2009, Simões et al., 2010), indicando a necessidade de cautela na interpretação dos resultados em grupos clínicos. Por outro, nem todos os índices são capazes de detectar os “simuladores” (cf. Franklin, 2002; Rosenfeld et al., 2010).

Congruente com estudos prévios, verificou-se um efeito da idade, no grupo da comunidade, no Ensaio 2 do TOMM (Teichner & Wagner, 2004), na FMD (Spar & La Rue, 2005), tempo total do *b-test* (Boone et al., 2000), bem como nos índices do 15-IMT (Simões et al., 2010). Por outro lado, os resultados refutam os dados de outros estudos, nomeadamente no efeito da idade encontrado neste grupo da comunidade no total de erros, tempo para contar pontos não agrupados e resultado total do DCT (cf. Boone et al., 2002a; Rudman et al., 2011). No que diz respeito à escolaridade, verifica-se, tal como no estudo de Boone e colaboradores (2002a), uma influência no número total de erros do DCT. A influência verificada no Ensaio 2 e no Ensaio de Reconhecimento do TOMM, no resultado total do *b-test*, bem como nos índices do 15-IMT não corroboram estudos prévios (cf. Toumbaugh, 1997; Teichner & Wagner, 2004; Boone et al., 2000; Simões et al., 2010). Apesar destas influências, verifica-se que na maioria dos índices os sujeitos continuam a ter desempenhos esperados (acima/abaixo dos pontos de corte definidos, consoante as medidas em análise).

Já no grupo clínico verifica-se um efeito da idade (cf. Boone et al., 2000) e escolaridade no resultado tempo total do *b-test*. A influência da escolaridade no total da MD e FMD poderá ser justificável pela influência da reserva cognitiva (Pires et al., 2010). Em relação ao DCT, verifica-se um efeito da idade na média de tempo para contar pontos agrupados, congruente com o estudo de Boone e colaboradores (2002) que, inversamente ao presente estudo, não verificaram este efeito também no resultado total. Apesar de se ter procedido à análise da influência de género, dada a não equivalência do número de sujeitos relativamente a esta variável nesta amostra, sugere-se a replicação destes estudos com amostras mais robustas.

Em relação ao grupo instruído para simular não se verifica influência da idade nos resultados. Ressalva-se que, como inicialmente este projecto seria para incluir indivíduos que estivessem à procura de obter reformas antecipadas, por invalidez ou incapacidade (sujeitos com eventual motivo para simular em contexto real), procurou-se que a amostra instruída a simular fosse o mais equivalente a esta (ou seja, com idades compreendidas entre os 55 e 65 anos) para que se pudesse proceder às comparações necessárias. Contudo, não foi dada autorização para recolha destes dados em tempo útil, e os sujeitos da amostra instruída já tinham sido avaliados, pelo que as idades do grupo instruído não são heterogéneas. Neste grupo instruído a simular, a escolaridade mostrou ter efeito significativo e moderado nos erros de comissão e no total do *b-test*, na razão do DCT, bem como no índice Vocabulário menos Memória de Dígitos. Estas influências da escolaridade poderão ser justificáveis pela teoria da reserva cognitiva que postula que a escolaridade e nível ocupacional têm um efeito protector (Pires et al., 2010). Por outro lado, foi possível observar que as mulheres “simulam” mais no Ensaio 2 do TOMM, no tempo para contar pontos agrupados, média de tempo dos pontos agrupados e total do DCT.

Em relação a variáveis afectivas, foi possível observar uma influência da depressão e ansiedade na pontuação total do SIMS, no grupo de controlo e DCL, o que corrobora a necessidade de ajuste dos pontos de corte deste instrumento para grupos clínicos específicos (Wisdom et al., 2010; Benge et al., *in press*). Os sujeitos da comunidade e com DCL com depressão/ansiedade pontuaram mais nas subescalas *perturbações afectivas*, *défi ce neurológico* e *perturbações mnésicas*. Considerando o ponto de corte

proposto por Wisdom e colaboradores (2010) (>24) verifica-se que os valores de sensibilidade continuariam adequados (89.66%) e os valores de especificidade elevados no grupo de controlo (96.77%), sendo que no grupo com DCL se verificaria uma melhoria significativa deste valor (94.12%), o que indica que o decréscimo dos valores de especificidade verificáveis no grupo clínico correspondem aos sujeitos deprimidos/ansiosos que estavam a ser classificados erradamente como “simuladores”. No grupo instruído, esta influência verifica-se no tempo total e média para contar pontos agrupados e valor da razão do DCT. Visto estas últimas medidas envolverem tempos de reacção, estas podem ser justificáveis pela diminuição da velocidade de processamento que pode ocorrer simultaneamente às alterações emocionais (Rudman et al., 2011). Contudo, esta influência pode dever-se também ao facto dos sujeitos estarem a fingir défices e, como tal, terem a percepção de que as perturbações emocionais causem lentificação das respostas, até porque, de acordo com os dados de Lee e colaboradores (2000), não se verificou relação entre depressão e sua severidade e resultados no DCT.

Todos estes resultados denotam que nem todos os TVS são, como hipotetizado, insensíveis a variáveis sociodemográficas, cognitivas e emocionais, o que reforça a necessidade do controlo destas variáveis e de estudos com grupos de idosos e em diferentes grupos clínicos (cf. Boone et al., 2000; Boone et al., 2002a; Swihart et al., 2008; Dean et al., 2009; Simões et al., 2010; Weiss & Rosenfeld, 2010; Strutt et al., 2011; Bengel, *in press*).

Analisando os valores de especificidade de cada uma das medidas, conclui-se que, tal como em estudos prévios os índices do TOMM (cf. Fernandes, 2009) e o índice Vocabulário menos Memória de Dígitos (cf. Dean et al., 2009) são insensíveis ao declínio. Os índices RMI, Pontuação total da MD, DCT (excepto valor da razão), Evocação Imediata (<6) do 15-IMT também continuam com taxas de falsos positivos reduzidas no grupo da comunidade e DCL e, como tal, são aparentemente insensíveis ao declínio cognitivo, não havendo estudos do nosso conhecimento realizados com estes grupos. Congruente com alguns estudos, o presente trabalho evidencia a sensibilidade de algumas medidas ao declínio cognitivo (FMD, 3 e 4 dígitos cronometrados em sentido directo, SIMS, *b-test*, razão do DCT e evocação imediata (<9) do 15IMT) e, conseqüentemente, a necessidade de ajuste dos pontos de corte (cf. Boone et al, 2002a; Teichner & Wagner, 2004; Dean et

al., 2009; Fernandes, 2009; Simões et al., 2010). Na amostra da comunidade, as taxas de falsos positivos são reduzidas na maioria das medidas (excepto erros de comissão do *b-test* e resultado combinado do 15-IMT). Por seu turno, verifica-se um pequeno decréscimo nos valores de especificidade na amostra de DCL na FMD, 3 e 4 dígitos cronometrados em sentido directo, SIMS, *b-test*, razão do DCT e evocação imediata (<9) do 15IMT). Não obstante estes valores, ressalva-se os baixos níveis de sensibilidade de todos os índices em estudo, excepto o SIMS. O SIMS é a medida com melhor capacidade para detectar correctamente os “simuladores” e com valores de falsos positivos reduzidos no grupo da comunidade. Tal como referido na literatura, os valores de falsos positivos reduzidos são preferíveis a altas taxas de sensibilidade, sob pena de se classificar erradamente os indivíduos como “simuladores”. Os baixos valores de sensibilidade poderão ser “compensados” com o recurso a diferentes medidas de exagero de sintomas, podendo-se deste modo tomar uma decisão mas fidedigna (Rosenfeld et al., 2010; Bliger, 2012). O RMI, TOMM, pontuação total da MD, Evocação Imediata do 15-IMT e Vocabulário menos Memória de Dígitos são os índices que são, de acordo com os dados deste estudo, insensíveis ao envelhecimento e declínio.

Foi possível observar que os indivíduos de facto não simulam todos nos mesmos domínios nem nos mesmos instrumentos, o que reforça a premissa relativa à pertinência do recurso a diferentes TVS, de forma a poder detectar estes sujeitos, até porque estes são selectivos nos domínios que escolhem simular (cognitivo, físico e/ou emocional) (cf. Tombaugh, 1997; Boone et al., 2000; Clegg et al., 2009; Rosenfeld et al., 2010; Simões et al., 2010; Wisdom et al., 2010; Rudman et al., 2011). Analisando estes dados verifica-se que, nos sujeitos instruídos a simular, o DCT não detecta os “simuladores” na maioria dos seus índices, observando-se que, os poucos sujeitos que simulavam défices em alguns índices deste instrumento eram também assinalados noutros TVS associados a valores de sensibilidade mais elevados. Como tal, parece existir evidência de que o DCT não acrescenta validade adicional na detecção de simulação no presente protocolo, contrariamente aos estudos realizados com o DCT que reportam valores de sensibilidade consideráveis (Boone et al., 2002a; Rudman et al., 2011). Mesmo após ajuste dos pontos de corte verificou-se que os índices deste

instrumento continuavam com valores de sensibilidade nulos, apesar dos elevados valores de especificidade. O que estes resultados permitem inferir é que, apesar de estarem construídos com base no pressuposto de efeito de chão (os itens são tão fáceis que é difícil sujeitos com défices reais errarem) (Tombaugh, 1997; Franklin, 2008), poderá ter-se caído no extremo oposto de efeito de tecto (os itens serem de tal forma exigentes que os sujeitos não são detectados nessas medidas).

O ponto de corte proposto por Fernandes (2009) para o Ensaio 1 do TOMM (<33) revelou-se o mais adequado como medida de *screening*; contudo, este ponto de corte poderá ser útil para se colocar a hipótese de simulação ou esforço insuficiente mas observou-se que sujeitos instruídos a simular também obtiveram pontuações superiores a este valor no ensaio 1 e ainda valores superiores a 45 no Ensaio 2 e no Ensaio de Reconhecimento do TOMM. Os resultados permitem concluir que, contrariamente a estudos prévios (cf. O’Byrant et al., 2008; Fernandes, 2009) resultados superiores a 45 no Ensaio 1 não implicam desempenhos semelhantes nos restantes ensaios e, conseqüentemente, a sua não aplicação. Tal como referido por Bliger (2012), os pontos de corte são necessários pois constituem um método de classificação mas serão mais úteis se forem encarados como *guidelines*, mais do que considerados como diferenciadores entre a presença/ausência de determinada característica.

Este estudo acarreta algumas limitações. Primeiramente refere-se a ausência de dados normativos no ACE-R por idade e escolaridade nos domínios cognitivos para os sujeitos com idades compreendidas entre os 55-64 anos.

Um outro limite do presente trabalho prende-se com o facto de não serem incluídos no presente estudo sujeitos com eventual motivo para “simulação/exagero de sintomas”. Apesar de inicialmente ter sido colocada a hipótese de avaliar sujeitos em contexto de pedidos de reforma antecipada/invalidez, tal cenário não pode ser alvo de análise. O objectivo seria a comparação de desempenhos entre o grupo em contexto real e o grupo instruído a simular, até para analisar empiricamente algumas críticas que surgem na literatura e que referem que os valores adequados de sensibilidade e especificidade elevados são obtidos com sujeitos instruídos a simular (Clegg et al., 2009; Greve, et al., 2009; Rosenfeld et al., 2010).

Uma outra limitação remete para o tamanho das amostras em comparação e a não equivalência entre as mesmas. Em trabalhos futuros, poderá ser útil alargar as amostras de adultos idosos, de forma a obter resultados mais robustos (com especial ênfase na amostra clínica, procurando avaliar sujeitos mais escolarizados e mesmo analfabetos) além de tornar possível a análise do contributo de cada medida através de modelos de regressão.

Refere-se ainda a extensão do presente protocolo que, em contexto hospitalar e com sujeitos com algum declínio, torna difícil a sua concretização. A aplicação dos diferentes instrumentos teve de ser repartida constituindo foco de maior resistência na participação no presente projecto.

Face à reduzida validade do DCT no presente estudo poderá ser útil em estudos futuros proceder à análise do *21-Item Test* como instrumento a incluir neste protocolo. Este instrumento consiste em dois ensaios em que inicialmente os sujeitos têm de evocar uma lista de 21 palavras previamente lidas (ensaio de evocação livre); no ensaio seguinte, são lidos aos sujeitos 21 pares de palavras e estes têm de identificar a palavra que pertencia à lista inicial. Ryan, Turpin & Kreiner (2012) analisaram a capacidade deste teste na detecção de simulação ou esforço insuficiente na população idosa verificando que o ensaio de escolha forçada é insensível à idade e também relativamente insensível ao declínio cognitivo, sendo aparentemente útil e indicado para utilizar com a população idosa. Os autores observaram que 100% dos sujeitos idosos da comunidade foram classificados correctamente como “honestos” e que este instrumento tem valores de especificidade elevados, mesmo em amostras de sujeitos com problemas cognitivos.

Surge ainda a necessidade de aprofundar o estudo das diferenças nos desempenhos dos diferentes tipos de DCL (amnésico *versus* não amnésico) em amostras mais representativas, de forma a fornecer dados mais sólidos e robustos, bem como em sujeitos com diagnóstico de demência para estudar as medidas mais adequadas nesta população.

VI - Conclusões

Os resultados nas avaliações neuropsicológicas podem não traduzir as competências reais dos sujeitos. Estes podem exagerar sintomas, em

circunstâncias em que existem ganhos secundários que podem advir desses comportamentos. Na população idosa existem também circunstâncias em que os indivíduos podem exagerar sintomas, de forma a obter algum tipo de recompensa externa, como por exemplo nos casos de reformas antecipadas, casos de erros médicos de modo a obter recompensa monetária, ou mesmo determinado tratamento. Nestes contextos, torna-se essencial o recurso a TVS. Por questões diversas, estes instrumentos nem sempre são incluídos nas baterias de avaliação neuropsicológica mas a evidência empírica da importância do seu uso tem vindo a ser documentada na literatura.

No presente trabalho foi possível corroborar a utilidade de recorrer a diferentes TVS, facultando cada um o seu contributo específico para identificação de simulação/exagero de sintomas (cf. Rosenfeld et al., 2010). No entanto, pode surgir a possibilidade de obtenção de resultados contraditórios nas diferentes medidas. Por isso, os resultados não devem ser considerados como evidência definitiva de presença/ausência de simulação ou esforço insuficiente. Tal como evidenciado na literatura, há que conjugar informação proveniente de diferentes fontes articulando os resultados dos TVS com outros dados disponíveis, como avaliações neuropsicológicas efectuadas, dados de neuroimagem, observação dos comportamentos, auto relatos, registos clínicos (cf. Clegg et al., 2009; Yochim et al., 2010; Simões et al., 2010; Rosenfeld et al., 2010), não descurando ainda a prática e perícia dos clínicos.

Se os TVS forem usados de forma dicotómica no sentido de categorizar os desempenhos dos sujeitos como válidos/inválidos, torna-se inevitável a ocorrência de falsos positivos ou negativos. Não pode ser descurado que sujeitos com determinadas condições neurológicas e psiquiátricas podem “fracassar” nestas medidas, por etiologias que não se devam aos desempenhos inválidos (Bliger, 2012).

Por último emerge novamente a evidência da necessidade de ajuste dos pontos de corte para os grupos clínicos, sob pena de classificar erradamente estes sujeitos como estando a “simular” ou a “exagerar sintomas”.

Bibliografia

Albert, M., DeKosky, S., Dickson, D., Dubois, B., Feldman, H., Fox, N.,

- Gamst, A., Holtzman, D., Jagust, W., Petersen, R., Snyder, P., Carrillo, M., Thies, B., & Phelps, C. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup. *Alzheimer's & Dementia*, 7(3), 270-279. doi: 10.1016/j.jalz.2011.03.008
- American Psychiatric Association (APA, 2002). *DSM-IV-TR: Manual de diagnóstico e estatística das perturbações mentais* (4ªed., texto revisto). Lisboa: Climepsi Editores.
- Ashendorf, L., Constantinou, M., & McCaffrey, R. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology*, 19, 125-130.
- Bauer, L., O'Bryant, S., Lynch, J., McCaffrey, R., & Fisher, J. (2007). Examining the Test of Memory Malingering Trial 1 and Word Memory Test Immediate Recognition as screening tools for insufficient effort. *Assessment*, 14(3), 215-222. doi: 10.1177/1073191106297617
- Benge, J., Wisdom, N., Collins, R., Franks, R., LeMaire, A., Chen, D. (in press). Diagnostic utility of the Structured Inventory of Malingered Symptomatology for identifying psychogenic non-epileptic events. *Epilepsy & Behavior*. doi: 10.1016/j.yebeh.2012.05.007
- Blinger, E. (2012). Symptom validity testing, Effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 1-11. doi: 10.1017/S1355617712000252
- Boone, K., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., Warner-Chacon, K., & Berman, N. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b Test. *Archives of Clinical Neuropsychology*, 15(3), 227-241.
- Boone, K., Lu, P., Back, C., King, C., Lee, A., Phipott, L., Shamieh, E., & Warner-Chacon, K., (2002a). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology*, 17, 625-642.
- Boone, K., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002b). The Rey 15-Item Recognition Trial: A technique to enhance sensitivity of the Rey 15-Item Memorization Test. *Journal of*

- Clinical and Experimental Neuropsychology*, 24(5), 561-573.
- Bush, S., Ruff, R., Tröster, A., Barth, J., Koffler, S., Pliskin, N., Reynolds, C., & Silver, C. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, 20, 419-426. doi: 10.1016/j.acn.2005.02.002
- Clegg, C., Fremouw, W., & Mogge, N. (2009). Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among outpatients seeking to claim disability. *The Journal of Forensic Psychiatry & Psychology*, 20(2), 239-254. doi: 10.1080/14789940802267760
- Dean, A., Victor, T., Boone, K., Philpott, L., & Hess, R. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, 23(1), 133-152. doi: 10.1080/13854040701819050
- Denning, J. (2012). The efficiency and accuracy of the Test of Memory Malingered Trial 1, errors on the first 10 items of the Test of Memory Malingered, and five embedded measures um predicting invalid test performance. *Archives of clinical Neuropsychology*, 27, 417-432. doi: 10.1093/arclin/acs044
- Fernandes, S. (2009). Test of Memory Malingered (TOMM): Estudos de validação em adultos idosos com Declínio Cognitivo Ligeiro. Dissertação de Mestrado não publicada. Faculdade de Psicologia e Ciências da Educação, Universidade de Coimbra.
- Franklin, K. (2008). Malingering as a dichotomous variable: Case report on an insanity defendant. *Journal of Forensic Psychology Practice*, 8(1), 95-107. doi: 10.1080/15228930801949241
- Greve, K., & Bianchini, K. (2006). Should the retention trial os the Test of Memory Malingered be optional?. *Archives of Clinical Neuropsychology*, 21, 117-119. doi: 10.1016/j.acn.2005.06.009
- Greve, K., Springer, S., Bianchini, K., Black, F., Heinly, M., Love, J., Swift, D., & Ciota, M. (2007). Malingering in toxic exposure: Classification accuracy of Reliable Digit Span and WAIS-III Digit Span scaled score. *Assessment*, 14(1), 12-21. doi: 10.1177/1073191106295095

- Greve, K., Etherton, J., Ord, J., Bianchini, K., Curtis, K. (2009). Detecting malingering pain-related disability: Classification accuracy of the Test of Memory Malingering. *The Clinical Neuropsychologist*, *23*, 1250-1271. doi: 10.1080/13854040902828272
- Hilsabeck, R., Gordon, S., Hietpas-Wilson, T., & Zartman, A. (2011). Use of trial 1 of the Test of Memory Malingering (TOMM) as a screening measure of effort: Suggested discontinuation rules. *The Clinical Neuropsychologist*, *1*, 1-11. doi: 10.1080/13854046.2011.589409
- Iverson, G. (2003). Detecting malingering in civil forensic evaluations. In A. Horton & L. Hartlage (Eds.), *Handbook of forensic neuropsychology* (pp. 137-177). New York: Springer Publishing Company.
- Jasinski, L., Berry, D., Shandera, A., & Clark, J. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 300-314. doi: 10.1080/13803395.2010.516743
- Jelicic, M., Ceunen, E., Peters, M., Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, *67*(9), 850-855. doi: 1.1002/jclp.20805
- Kizilbash, A., Vanderploeg, R., & Custiss, G. (2002). The effects of depression and anxiety on memory performance. *Archives of Clinical Neuropsychology*, *17*, 57-67.
- Larrabee, G. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 1-7. doi: 10.1017/S1355617712000240
- Lee, G., Loring, D., & Martin, R. (1992). Rey's 15-Item Visual Memory Test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment*, *4*(1), 43-46.
- Lee, A., Boone, K., Lesser, I., Wohl, M., Wilkins, S., & Parks, C. (2000). Performance of older depressed patients on two cognitive malingering tests: False positive rates for the Rey 15-item

- Memorization and Dot Counting Tests. *The Clinical Neuropsychologist*, 14(3), 303-308.
- Lu, Po H., Rogers, S., & Boone, K. (2007). Use of standart memory tests to detect suspect effort. In K. Boone (Ed), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 128-151). New York: The Guilford Press.
- O'Bryant, S., & Lucas, J. (2006). Estimating the predictive value of the Test of Memory Malinger: An illustrative example for clinicians. *The Clinical Neuropsychologist*, 20, 533-540. doi: 10.1080/13854040590967568
- O'Bryant, S., Engle, L., Kleiner, J., Vasterling, J., & Black, F. (2007). Test of Memory Malinger (TOMM) Trial 1 as a screening measure for insufficient effort. *The Clinical Neuropsychologist*, 21, 511-521. doi: 10.1080/13854040600611368
- O'Bryant, S. E., Finlay, C., & O'Jile, J. R. (2007). TOMM performance and self-reported symptoms of depression and anxiety. *Journal of Psychopathology and Behavioral Assessment*, 29, 111-114. doi: 10.1007/s10862-006-9034-9
- O'Bryant, S., Gavett, B., McCaffrey, R., O'Jile, J., Huerkamp, J., Smitherman, T. (2008). Clinical utility of trial 1 of the Test of Memory Malinger (TOMM). *Applied Neuropsychology*, 15, 113-116. doi: 10.1080/09084280802083921
- Pires, L.M., Simões, M. R., & Firmino, H. (2010). Envelhecimento e doença de Alzheimer: Implicações para a hipótese da reserva cognitiva. *Psicologia, Educação e Cultura*, 14(1), 5-22.
- Rees, L., Tombaugh, T., Gansler, D., Moczynski, N. (1998). Five validation experiments of the Test of Memory Malinger (TOMM). *Psychological Assessment*, 10(1), 10-20.
- Rees, L., Tombaugh, T., & Boulay, L. (2001). Depression and the Test of Memory Malinger. *Archives of Clinical Neuropsychology*, 16, 501-506.
- Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21(3), 182-188.
- Rogers, R., Sewell, K., & Goldstein, A. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior*,

- 18(5), 543-552.
- Rogers, R., & Granacher, R. (2011). Conceptualization and assessment of malingering. In E. Drogin, F. Dattilio, R. Sadoff, & T. Gutheil (Eds), *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 659-678). New Jersey: John Wiley & Sons, Inc.
- Rosenfeld, B., Green, D., Pivovarova, E., & Dole, T. (2010). What to do with contradictory data? Approaches to the integration of multiple malingering measures. *International Journal of Forensic Mental Health, 9*, 63-73. doi: 10.1080/14999013.2010.499559
- Rudman, N., Oyebode, J., Jones, C., & Bentham, P. (2011). An investigation into validity of effort tests in a working age dementia population. *Aging & Mental Health, 15*(1), 47-57. doi: 10.1080/13607863.2010.508770
- Ryan, J., Turpin, D., & Kreiner, D. (2012). Specificity of the 21-Item Test in two elderly samples. *Aging, Neuropsychology, and Cognition, 1*, 1-18. doi: 10.1080/13825585.2011.646941
- Simões, M. R. (2005). O exame dos comportamentos de simulação em avaliação (neuro)psicológica. In C. Vieira, A. M. Seixas, A. Matos, M. P. Lima, M. Vilar & M. R. Pinheiro (Eds.), *Ensaio sobre o comportamento humano: Do diagnóstico à intervenção. Contributos nacionais e Internacionais* (pp. 453-481). Coimbra: Almedina.
- Simões, M. R. (2006). Testes de validade de sintomas na avaliação de comportamentos de simulação. In A. C. Fonseca, M. R. Simões, M. C. T. Simões e M. S. Pinho (Eds.), *Psicologia forense* (pp. 279-309). Coimbra: Almedina.
- Simões, M. R., Sousa, L., Duarte, P., Firmino, H., Pinho, M., Gaspar, N., Pires, L., Batista, P., Silva, A., Silva, S., Ferreira, A., & França, S. (2010). Avaliação da simulação ou esforço insuficiente com o Rey 15-Item Memory Test (15-IMT): Estudos de validação em grupos de adultos idosos. *Análise Psicológica, 1*(XXVIII), 209-226.
- Slick, D. J., Sherman, E. M. S. & Iverson, G. L. (1999). Diagnostic criteria for Malingered Neurocognitive Dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 4*, 545-561.
- Spar, J., & La Rue, A. (2005). Envelhecimento Normal. In J. Spar & A. La

- Rue (Eds.), *Guia prático de Psiquiatria Geriátrica* (pp. 35-70). Lisboa: Climepsi.
- Strutt, A., Scott, B., Shrestha, S., & York, M. (2011). The Rey 15-Item Memory Test and spanish-speaking older adults. *The Clinical Neuropsychologist*, 25(7), 1253-1265. doi: 10.1080/13854046.2011.609839
- Swihart, A., Harris, K., Hatcher, L. (2008). Inability of the Rarely Missed Index to identify simulated malingering under more realistic assessment conditions. *Journal of Clinical and Experimental Neuropsychology*, 30(1), 120-126. doi: 10.1080/13803390701249044
- Teichner, G., & Wagner, M. (2004). The test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, 19, 455-464. doi: 10.1016/S0887-6177(03)00078-7
- Tombaugh, T. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260-268.
- Weiss, R., & Rosenfeld, B. (2010). Cross-Cultural validity in malingering assessment: The Dot Counting Test in a rural indian sample. *International Journal of Forensic Mental Health*, 9, 300-307. doi: 10.1080/14999013.2010.526680
- Wisdom, N., Callahan, J., & Shaw, T. (2010). Diagnostic utility of the Structured Inventory of Malingered Symptomatology to detect malingering in a forensic sample. *Archives of Clinical Neuropsychology*, 25(2), 118-125.
- Yanez, Y., Fremouw, W., Tennant, J., Strunk, J., Coker, K. (2006). Effects of severe depression on TOMM performance among disability-seeking outpatients. *Archives of Clinical Neuropsychology*, 21, 161-165. doi: 10.1016/j.acn.2005.07.009
- Yochim, B., Kane, K., Horning, S., & Pepin, R. (2010). Malingering or expected deficits? A case of herpes simplex encephalitis. *Neurocase*, 16(5), 451-460. doi: 10.1080/13554791003623334

Anexos

**Anexo 1: Critérios de diagnóstico de Simulação de Perturbação Neurocognitiva (SPN)
(Slick et al., 1999)**

- A) Presença de incentivos externos;
- B) Evidências na avaliação neuropsicológica
 - 1. Enviesamento definitivo das respostas
 - 2. Enviesamento provável
 - 3. Discrepância entre os dados obtidos nos testes e padrões de funcionamento cognitivo conhecidos
 - 4. Discrepância entre os dados obtidos nos testes e os dados observacionais;
 - 5. Discrepância entre os dados obtidos nos testes e dados de outros informadores fidedignos
 - 6. Discrepância entre os dados obtidos nos testes e dados de registos do sujeito
- C) Evidências de exagero no auto-relato do sujeito;
 - 1. Discrepância entre auto relato e história documental
 - 2. Sintomas auto repostados discrepantes com padrões conhecidos de funcionamento cognitivo
 - 3. Discrepância entre auto-relato e comportamento observado
 - 4. Discrepância entre sintomas referenciados pelo próprio e informação obtida com outros informadores
 - 5. Evidência de falsificação/exagero de disfunção psicológica
- D) Estes comportamentos não podem ser melhor explicados por doença psiquiátrica, neurológica ou desenvolvimental.

Segundo Slick e colaboradores (1999) a Simulação de Perturbação Neurocognitiva (SPN) pode ser: Definitiva (Critério A, critério B1 e critério D); Provável [Critério A; dois ou mais critérios B (excluindo B1) ou um critério B (excluindo B1) e um ou mais critérios C; e Critério D]; ou possível [(Critério A; um ou mais critérios C; critério D); ou (todos os critérios para SPN provável ou definitiva mas sem preencher o critério D)].

Anexo 2: Itens do Índice de Respostas Infrequentes e respectiva ponderação

Table 3. LMDR items and point values for the Rarely Missed Index.

Item	Question	Point Value
12	Was the rent due?	-22
16	Was the man's name Joe Garcia?	55
18	Was it 6:00?	84
22	Was the program interrupted?	67
24	Was the storm expected to stay in the area through the night?	13
29	Did Joe decide to stay home?	7

Note: LMDR = WMS-III Logical Memory Delayed Recognition task; RMI point values as presented by Killgore and DellaPietra (2000).

Anexo 3: Instrução dada à amostra da comunidade instruída a simular

"Procure colocar-se na seguinte situação: IMAGINE que teve um acidente que lhe causou dificuldades de memória e que, por esse motivo, solicita à Segurança Social um pedido de pensão por incapacidade.

Como consequência destas dificuldades começou a sentir-se mais triste, perdeu interesse em actividades que anteriormente gostava de fazer, isolou-se, e passou a sentir dificuldades de concentração e em dormir.

Actualmente, estes sintomas e dificuldades **já não lhe acarretam dificuldades**, tendo o senhor/a senhora recuperado totalmente.

No entanto, como o(a) senhor(a) **pretende obter algum benefício (ganho ou recompensa) com o acidente que teve**, e apesar de já não ter problemas, vai **fingir que estes problemas e dificuldades estão ainda presentes**, e que o(a) impedem de trabalhar. **Esta situação poderá dar-lhe ainda acesso à reforma antecipada.**

A **Junta Médica da Segurança Social** onde se dirigiu encaminhou-o(a) a um psicólogo para avaliar a gravidade das suas queixas e dificuldades e, para isso, necessita de responder a testes psicológicos. **Os resultados nos testes psicológicos ajudarão a determinar se irá receber algum tipo de pensão ou reforma antecipada e a quantia que irá receber.**

Lembre-se que apresentar apenas sintomas não é suficiente; estes têm que mostrar que prejudicam a sua capacidade para trabalhar.

Contudo, deve fingir as suas dificuldades de forma a que sejam credíveis, porque, se for descoberto que está a fingir, o seu pedido será recusado, não receberá nada e terá de ir trabalhar.

Tenha em atenção, e não esqueça, que é muito importante convencer os médicos/profissionais que está realmente incapacitado(a) para trabalhar de modo a atingir o seu objectivo.

Tenha em atenção que poderão existir algumas perguntas específicas para detectar as pessoas que tentam fingir. "

Anexo 4: Análise comparativa do desempenho no grupo com DCL amnésico e não amnésico

Índices	
RMI	U=25,500; p=.364
TOMM	
Ensaio 2	U= 22,500, p=.230
Reconhecimento	U= 30,000, p=.669
Memória de Dígitos (MD)	
Pontuação Total	U= 28,000, p=.536
FMD	U= 27,000, p=.475
Tempo 3 dígitos	U= 18,000, p=.109
Tempo 4 dígitos	U= 25,000, p=.864
SIMS Total	U= 31,000, p=.740
b-test	
Tempo total resposta	U= 34,000, p=.962
Total erros comissão	U= 26,000, p=.417
Total erros omissão	U= 34,000, p=.962
Resultado Total (E-score)	U= 32,000, p =.813
Dot Counting Test	
Total Erros	U= 27,500, p=.475
Tempo total pontos agrupados	U= 35,000, p=1,000
Tempo total pontos não agrupados	U= 32,000, p=.813
Média para contar pontos agrupados	U= 31,000, p=.740
Razão	U= 30,000, p=.669
Resultado total (E-Score)	U= 27,000, p=.475
Rey 15-Item Memory Test	
Evocação Imediata	U= 25,000, p=.364
Resultado Combinado	U= 30,500, p=.669
Vocabulário – MD	U= 34,000, p=.962

Anexo 5: Análise dos valores de correlação entre depressão e ansiedade e os resultados do SIMS (amostra da comunidade e DCL)

	GDS-30	GAI
SIMS Total	0.593**	0.649**
<i>Perturbações Afectivas</i>	0.570**	0.700**
<i>Psicose</i>	0.246	0.277
<i>Capacidade Intelectual Reduzida</i>	0.107	0.218
<i>Défice Neurológico</i>	0.520**	0.506**
<i>Perturbações Amnésicas</i>	0.531**	0.484**

**p<0.01

Anexo 6: Médias e desvios padrão dos desempenhos nas diferentes medidas nos grupos Comunidade e DCL

	Comunidade	DCL
RMI	179.19 (36.331) (Min=107; Máx=226)	129.47 (56.966) (Min=-15; Máx=226)
b-test		
Tempo	10.52 (3.16) (Min=6.38; Máx=19.09)	13.23 (4.82) (Min=6.20; Máx=22.02)
Comissão	1.58 (2.96) (Min=0; Máx=14)	8.82 (17.23) (Min=0; Máx=72)
Omissão	15.52 (14.72) (Min=0; Máx=58)	31.76 (18.164) (Min=8; Máx=68)
Total	89.92 (73.15) (Min=22; Máx=318)	226.42 (346.637) (Min=34; Máx=1517)
SIMS total	9.39 (5.457) (Min=1; Máx=29)	15.00 (5.601) (Min=7; Máx=25)
Memória de Dígitos		
Pontuação total	11.48 (2.528) (Min=7; Máx=17)	8.65 (2.234) (Min=5; Máx=14)
FMD	9.52 (2.047) (Min=6; Máx=15)	7.53 (1.841) (Min=5; Máx=12)
3 dígitos cronometrados	1.437 (0.569) (Min=65; Máx=3.37)	1.987 (0.805) (Min=0.97; Máx=4.09)
DCT		
Tempo total agrupados	14.052 (7.178) (Min=4.45; Máx=36.86)	25.375 (10.922) (Min=11.89; Máx=46.41)
Tempo total não agrupados	32.249 (10.412) (Min=18.08; Máx=60.58)	36.916 (9.454) (Min=10.64; Máx=51.62)
Média agrupados	2.347 (1.198) (Min=0.74; Máx=6.14)	4.090 (1.916) (Min=1.77; Máx=7.74)
Média não agrupados	5.406 (1.729) (Min=3.01; Máx=10.10)	6.2935 (1.230) (Min=3.99; Máx=8.60)
Razão	2.595 (0.973) (Min=1; Máx=5.50)	1.862 (0.8199) (Min=0.82; Máx=3.57)
Resultado Total	8.77 (3.074) (Min=5; Máx=17)	11.18 (3.127) (Min=6; Máx=16)
Rey 15-item Memory Test		
Evocação Imediata	11.68 (3.310) (Min=4; Máx=15)	8.94 (3.092) (Min=3; Máx=14)
Resultado Combinado	23.61 (6.412) (Min=10; Máx=30)	16.71 (5.764) (Min=6; Máx=25)